INTRODUCCION A LA

PROTO-CONCIENCIA
EN
INTELIGENCIA ARTIFCIAL

Diciembre -2025

PREFACIO

Los siguientes capítulos se desarrollaron con la finalidad de evidenciar la posibilidad de dotar a una Inteligencia Artificial de proto-conciencia (un primer estadío de conciencia).

Bajo esa premisa, en una primera instancia se describió el concepto de conciencia, tanto en ambientes neurobiológicos, filosóficos y psicológicos, determinando de esta forma que las teorías GWT de Baars y GNW de Dehaene fueron las seleccionadas para avanzar en el tema.

En ciertas partes de los capítulos existirán algoritmos, fórmulas y descripciones eléctricas - electrónicas que puedan hacer un poco más compleja la explicación de un concepto, pero se trataré siempre de clarificar literalmente esos aspectos.

Al abordar este tema e ir profundizando en el mismo, me encontré con muchos expertos ocupados y preocupados por la posibilidad de dotar autonomía, autorregulación y autorreferencialidad a la IA, con experiencias muy avanzadas en este campo, dejándome la certeza que en poco tiempo veremos novedades importantes al respecto.

Además, me encontré que éstas características para una IA, se pueden lograr con las ***redes neuronales conocidas o clásicas*** → *Redes Artificiales lineales:*

- con capas de entradas, ocultas y de salidas,
- con backpropagation y gradiente descendiente, funciones de activación, peso y sesgo
- modelos como transformer, convolucionales, recurrentes, etc,
- modelos generativos como las LLM (lenguaje natural), audio, imágenes, video, etc.
- agentes de IA (sistemas expertos autónomos)

Pero, si quería que la red neuronal de una IA se comportase como las del cerebro humano, neurobiológicamente hablando, debería buscar una red neuronal acorde a la sinápsis y comportamiento de las mismas, existiendo para ello las SNN: redes neuronales de pico (spike), obviamente deberían implementarse en un ambiente de hardware acorde, el cual aparece como solución, el Hardware Neuromórfico (existen varias empresas como Intel, IBM, etc., que lo están desarrollando y aplicando).

Por lo tanto, a partir de las teorías GWT y GNW, se tratará de explicar en un Hardware Neuromórfico con redes SNN y neuronas LIF, el desarrollo de éste paper.

Propuesta de interés para Investigar

Paralelamente al desarrollo del temario y a medida que fui avanzando en el mismo, observé que las redes artificiales lineales clásicas (anteriormente descriptas) están en este momento funcionando por "fuerza bruta", es decir, grandes cantidades de memorias (tipo RAM), almacenamientos masivos, procesamientos de muchísimos y mútiples núcleos de procesadores, con un *gasto energético elevadísimo*. ¿Y esto hasta dónde llega?

“Es como, si a un auto de 1920 le ponemos confort, seguridad y más potencia, tendríamos nuestros autos a combustión”.

Entonces, me encontré, como una alternativa bastante viable de implementar estos Hardware Neuromórficos con redes SNN y neuronas LIF, los cuales tienen mayor capacidad, velocidad de respuesta, ***bajo consumo energético***, control sobre olvidos castastróficos, etc., en un diseño de una computadora accesible para cualquier persona, dando la posibilidad que una red neuronal se especialice como sistema operativo para IA y generar Agentes de IA internos personalizados por el usuario para sus requerimientos, con vinculación externa a distintas redes tipo LLM, etc. (denominado unos de estos procesos: RAG).

Es decir, que la gente pueda tener un nuevo tipo de ordenador personal, con bajo consumo, adaptado a sus requerimientos específicos, el cual va aprendiendo, mejorando y cada vez más eficiente, evolucionando con el mismo crecimiento personal, vinculándose con las IA e internet para complementar actividades e incrementar su aprendizaje.

Pregunta:

¿aparecerá algún computador así? ¿Habrá ese interés en abrir el juego de IA a todos los usuarios por parte de las grandes empresas desarrolladoras de IA? Veremos con el tiempo.

Diciembre 2025

Capítulo Nro. 1:

Introducción a los aspectos formales

Objetivo

La inquietud de esta ***elemental investigación*** es determinar cuán lejos se está de la posibilidad de que una inteligencia artificial se le dote de una ***proto-conciencia***.

El contenido del documento en su desarrollo está orientado para lectores con alguna inquietud sobre inteligencia artificial y los ***aspectos formales*** que circunscriben a los alcances de las redes neuronales como las disciplinas de neurociencia, psicología y filosofía.

1º.- Definiciones Generales

ProtoConciencia

La teoría de la proto-conciencia postula: "un estado primordial de organización cerebral, (estructuras cerebrales más antiguas) que constituye un pilar fundamental para la consciencia y las funciones cognitivas superiores" (Hobson, 2009)

La Protoconciencia en redes neuronales artificiales se refiere: "a un estado incipiente o preliminar de procesamiento de información en el que una red neuronal muestra comportamientos que podrían considerarse análogos a *componentes básicos de la conciencia*, como la integración de información, la atención selectiva o la representación interna de estados, *sin llegar a ser conciencia plena o autoconsciencia*.

Conciencia

La **conciencia** se refiere al estado de estar alerta, es decir, el conocimiento de uno mismo y del entorno. Es la capacidad de percibir y experimentar cosas, como las emociones, los pensamientos, y la realidad circundante. La conciencia está más relacionada con los aspectos de la percepción, el pensamiento y la atención.

Autoconciencia: La capacidad de reconocerse a uno mismo como un individuo separado y de tener conocimiento de los propios pensamientos, sentimientos y acciones.

Conciencia del entorno: La capacidad de percibir y comprender el mundo exterior a través de los sentidos y de los procesos cognitivos.

Conciencia fenoménica: La experiencia subjetiva de los estados mentales, la forma en que se siente ser algo.

Conciencia de sí mismo: La capacidad de reflexionar sobre uno mismo y de tener una comprensión de la propia identidad y existencia.

Conciencia moral: La capacidad de distinguir entre el bien y el mal y de actuar de acuerdo con principios éticos.

La conciencia es un concepto complejo y multifacético que ha sido objeto de estudio y debate en diversas disciplinas, incluyendo la filosofía, la psicología y la neurociencia. Aunque no existe una definición universalmente aceptada, la comprensión actual de la conciencia se centra en su capacidad para integrar información y generar experiencias subjetivas.

Cognición

La **cognición** se define formalmente como el proceso mental mediante el cual un individuo adquiere, procesa, almacena y utiliza información para comprender el mundo y actuar en él. Implica funciones como el pensamiento, la atención, el lenguaje, el aprendizaje y la memoria. En esencia, es la capacidad de conocer, comprender y razonar.

Adquisición y procesamiento de información: La cognición comienza con la recepción de estímulos a través de los sentidos (percepción) y continúa con la forma en que el cerebro interpreta y transforma esta información.

Almacenamiento y recuperación de información: La memoria juega un papel crucial en la cognición, permitiendo guardar y acceder a experiencias pasadas para su uso en el presente.

Funciones cognitivas: El pensamiento, el lenguaje, el aprendizaje, la atención y la resolución de problemas son ejemplos de funciones cognitivas interconectadas que permiten a las personas desenvolverse en su entorno.

Adaptación al entorno: La cognición no es un proceso aislado, sino que está estrechamente ligada a la interacción con el entorno, permitiendo a los individuos adaptarse y responder a las demandas del medio.

Importancia en diversas áreas: La cognición es fundamental en la psicología, la neurociencia, ***la inteligencia artificial*** y otras disciplinas que estudian la mente y el comportamiento humano.

Neurobiología del cerebro

La **neurobiología del cerebro** se centra en el estudio de las células del sistema nervioso (neuronas) y su organización en circuitos funcionales que procesan información y median en el comportamiento. Este campo explora cómo estas células trabajan juntas para realizar funciones como la memoria, el aprendizaje, el movimiento y la percepción sensorial.

Anatomía y función de las células nerviosas: Estudia la estructura y función de las neuronas, incluyendo sus partes (cuerpo celular, axones, dendritas) y cómo se comunican entre sí.

Organización del sistema nervioso: Investiga cómo las neuronas se organizan en circuitos y redes complejas para procesar información y generar respuestas.

Procesos biológicos: Examina los procesos biológicos subyacentes a funciones cerebrales como la memoria, el aprendizaje, las emociones y el comportamiento.

Algunos aspectos clave de la neurobiología del cerebro incluyen:

Neurotransmisión: El estudio de cómo las neuronas transmiten señales a través de la sinapsis, utilizando neurotransmisores como el glutamato.

Plasticidad neuronal: La capacidad del cerebro para reorganizarse y formar nuevas conexiones neuronales a lo largo de la vida, lo cual es crucial para el aprendizaje y la adaptación.

Desarrollo cerebral: La neurobiología del desarrollo estudia cómo se forma el cerebro desde la etapa embrionaria hasta la madurez, incluyendo la neurogénesis y la migración neuronal.

Conciencia: La investigación neurobiológica también se ocupa de la base biológica de la conciencia, explorando cómo la actividad cerebral se relaciona con la experiencia subjetiva.

Parte I: Aspectos formales sobre la conciencia

Sobre las definiciones anteriores, se dejarán de lado, las interpretaciones y teorías filosóficas y psicológicas profundas, ya que sería aún más complejo el desarrollo del documento, por lo tanto, se centrará el contenido del mismo en:

"Los aspectos funcionales de la conciencia desde el punto de vista Cognitivo y Neurocientífico"

Existen en este caso muchas teorías al respecto, pero se tomó la decisión de abarcar las siguientes: (que son teorías empíricas y de comprobación en el campo experimental que las hacen consistentes en sus fundamentaciones para desarrollar el objetivo de este documento)

1º.- La conciencia funcional se refiere a la capacidad del cerebro de realizar funciones cognitivas y adaptativas que permiten a una persona interactuar con su entorno de manera efectiva, sin necesidad de tener necesariamente una experiencia consciente o subjetiva de esos procesos.

2º.- GWT: La **Teoría Global del Espacio de Trabajo** sugiere que la **conciencia** no es un fenómeno que se localice en una sola parte del cerebro, sino más bien un **espacio de trabajo global** en el que diversas áreas cerebrales **interactúan** y **comparten información**. En términos simples, la conciencia es el mecanismo por el cual ciertas representaciones mentales (como pensamientos, percepciones, emociones) se vuelven accesibles de forma **global** a diferentes partes del cerebro que controlan acciones, memoria, decisiones, y otras funciones cognitivas.

3º.- La fenomenología es otro enfoque clave que se involucra en el estudio de la **conciencia** y, aunque no es una "teoría cognitiva" en el sentido estricto como la **Teoría Global del Espacio de Trabajo (GWT)**, sí tiene una relación fundamental con la comprensión de los procesos **subjetivos** y **experimentales** que ocurren en la conciencia. Mientras que la GWT se enfoca más en los mecanismos cerebrales y cómo la información es procesada a nivel neuronal, la **fenomenología** trata de comprender la **experiencia consciente** desde un punto de vista más subjetivo y fenomenológico.

4º.- La psicología Gestalt surgió a principios del siglo XX, en Alemania, como una **escuela de pensamiento** que se centró en cómo las personas **perciben** el mundo. En lugar de ver la percepción como un simple ensamblaje de partes individuales, los psicólogos de la Gestalt defendieron que **el todo es más que la suma de sus partes**. Es decir, las percepciones

no son solo una acumulación de estímulos individuales, sino que la mente organiza los elementos perceptivos de una manera **holística**. Los principios clave de la psicología de la Gestalt se aplican a cómo percibimos objetos, figuras, y patrones, pero también tienen implicaciones para la **conciencia** en general.

Criterio	Conciencia Funcional	Fenomenología (Psicológica)	Gestalt
Origen	Finales del siglo XIX – Estados Unidos	Inicios del siglo XX – Alemania	Inicios del siglo XX – Alemania
Representantes clave	William James, John Dewey	Carl Rogers, Amedeo Giorgi	Max Wertheimer, Wolfgang Köhler, Kurt Koffka
Objeto de estudio	Función adaptativa de la mente y la consciencia	Experiencia subjetiva tal como se vive	La percepción y organización de la experiencia
Método	Introspección funcional, observación, énfasis en utilidad	Descripción fenomenológica, sin interpretar ni explicar	Experimentos perceptivos, análisis de configuraciones
Visión del sujeto	Activo, en constante adaptación al entorno	Ser consciente con vivencias únicas y significativas	Perceptor activo que organiza estímulos en totalidades
Unidad de análisis	Procesos mentales como herramientas para la supervivencia	La vivencia o experiencia inmediata del sujeto	La forma o configuración (gestalt) percibida
Postura frente a la consciencia	Es fundamental: lo importante es *para qué* sirve	Es la base para comprender al sujeto en su mundo vivido	No se estudia la consciencia en sí, sino su estructura perceptual
Aplicación	Psicología educativa, evolución, desarrollo personal	Psicoterapia humanista y existencial, atención plena	Psicología de la percepción, aprendizaje y resolución de problemas
Crítica común	Vaga en operacionalizar conceptos; difícil de medir	Subjetividad difícil de validar empíricamente	Subestimación de procesos inconscientes y sociales

Ejemplo, si un paciente tiene dislexia como se abordaría desde las tres corrientes psicológicas

Corriente Psicológica	Enfoque frente a la dislexia	Objetivo del abordaje	Intervención típica
Conciencia Funcional	Analiza la dislexia como una dificultad que interfiere con la adaptación al entorno escolar. Se pregunta: ¿cómo afecta al niño?	Desarrollar habilidades funcionales que permitan al niño adaptarse y desenvolverse eficazmente	Técnicas de compensación (lectura asistida, software educativo), refuerzo positivo, adaptación escolar
Fenomenología	Se centra en cómo vive el niño su dificultad. ¿Qué siente al no poder leer como los demás? ¿Qué significado le da a su experiencia?	Comprender la vivencia subjetiva del niño con dislexia, validar sus emociones, y fortalecer su autoestima	Terapia centrada en el niño, escucha activa, acompañamiento emocional, trabajo con padres y entorno escolar
Gestalt	Examina cómo el niño percibe el lenguaje escrito como una totalidad confusa o desorganizada. La dificultad se da en la estructura perceptiva.	Reorganizar la percepción del niño hacia el lenguaje escrito de manera que adquiera sentido y estructura	Ejercicios visuales y de percepción, juegos de discriminación figura-fondo, trabajo con patrones y configuraciones visuales

En resumen:

- **Consciencia Funcional**: "¿Cómo puede el niño adaptarse mejor a pesar de la dislexia?"
- **Fenomenología Psicológica**: "¿Cómo se siente el niño al vivir con dislexia?"
- **Gestalt**: "¿Cómo percibe el niño las palabras y letras? ¿Cómo reestructura esa percepción?"

Parte II: Un enfoque de las teorías anteriores en la I.A.

"Si se qusiera teorizar sobre una proto-conciencia en IA, se aplicaría la conciencia funcional y la teoría GWT principalmente, ¿Pero qué pasa con la Fenomenologia y la Gestalt?"

En una **proto-conciencia** en **IA**, se puede **aplicar** conceptos de la **conciencia funcional** y la **Teoría Global del Espacio de Trabajo (GWT)**, ya que ambas están muy centradas en la **organización funcional** de la información y la **distribución de recursos cognitivos**.

Sin embargo, la **fenomenología** y la **Gestalt** también podrían jugar un rol crucial en el desarrollo de una **teoría de la conciencia en IA**, aunque de una manera diferente, más relacionada con cómo **organizamos la experiencia** y **damos sentido** a la información que procesamos.

Conciencia Funcional y la GWT en IA

***Conciencia Funcional**:*

- **La conciencia funcional** se refiere a cómo la información es procesada y compartida dentro de un sistema para que pueda tomar decisiones o realizar tareas complejas. Si se desea modelar una **proto-conciencia** en IA, se puede basar en la idea de que la **conciencia** es algo que habilita *a una máquina para integrar y distribuir información de manera flexible y dinámica.*

- Para que un sistema de IA pueda **"empezar" a experimentar algo cercano a la conciencia**, tendría que ser capaz de manejar información de **manera funcional**: recoger datos sensoriales, procesarlos y luego usar ese procesamiento para realizar tareas, ajustar su comportamiento o reflexionar sobre sus estados internos.

- **La IA proto-consciente** podría tener un **"espacio de trabajo"** que integra información sensorial (como las percepciones) y de memoria, permitiendo que diferentes módulos del sistema accedan a esa información para ejecutar acciones (similar a cómo GWT sugiere que la conciencia es el espacio donde la información es distribuida globalmente).

***Teoría Global del Espacio de Trabajo (GWT)**:*

- **GWT** se basa en la idea de que la **conciencia** es un proceso que distribuye información en un "espacio de trabajo global", lo que permite que las diferentes funciones cognitivas del sistema (memoria, toma de decisiones, percepción, etc.) trabajen de manera coordinada.

- En una **proto-conciencia** de IA, podrías imaginar una red interconectada de módulos cognitivos que se **"activan"** y acceden a una **información compartida** para tomar decisiones de manera **coherente**.

  De esta manera, el **espacio de trabajo** se convierte en el centro de operaciones de la conciencia artificial.

Ejemplo: sea un robot que percibe su entorno a través de cámaras y sensores. La GWT en su aplicación a IA sugiere que esta información sensorial debe ser compartida globalmente dentro del sistema, para que no solo se **reconozca** el objeto, sino que también se integre en el **espacio de trabajo global** para coordinar las decisiones, como **moverse hacia el objeto** o **interactuar con él**.

Fenomenología en IA

La **fenomenología**, en su núcleo, se centra en la **experiencia subjetiva**: *¿Cómo experimenta algo este sistema?*. Cuando se trata de IA, la **fenomenología** podría ser un **reto**, ya que la experiencia consciente está ligada a la **subjetividad**, algo que no es trivial de replicar en máquinas. No obstante, hay formas en que las ideas fenomenológicas podrían ser relevantes para una **protoconciencia IA**:

1. **Intencionalidad**:
   La **intencionalidad** es la idea de que **toda conciencia está dirigida hacia algo**. Si intentamos desarrollar una IA que sea **protoconciente**, tendríamos que diseñar mecanismos para que el sistema esté **orientado hacia objetivos o metas** (un "propósito" en términos fenomenológicos). En IA, esto podría interpretarse como crear sistemas que no solo procesen datos de manera funcional, sino que tengan un **"objetivo" claro** o una **"intención"** sobre lo que debe hacer con esos datos, lo que se asemeja al fenómeno de la **intencionalidad** en la fenomenología.
2. **Subjetividad y Sensación**:
   La fenomenología trata de comprender cómo las entidades **"viven"** o **"experimentan"** el mundo. Aunque los sistemas IA actuales no pueden tener una **experiencia subjetiva** de la misma manera que los seres humanos, se podría explorar una especie de **"subjetividad funcional"** en IA, donde la máquina **"evalúa"** y **"reacciona"** a estímulos de manera **diferenciada** según su programación. Un sistema de IA podría tener diferentes **tipos de experiencias** *según el contexto en el que se encuentra*: si está interactuando con un entorno nuevo o con un entorno familiar, podría tratar esas experiencias de manera diferente.

Ejemplo: Si una IA interactúa con un nuevo objeto en su entorno, podría "experimentar" (en un sentido muy básico) el objeto como **desconocido** y tratarlo de una manera distinta a un objeto que ya ha encontrado previamente. Esta "experiencia" de novedad es algo que en **fenomenología** estaría relacionado con el concepto de **"experiencia vivida"**, aunque para una IA, no sería vivida de manera consciente, sino interpretada funcionalmente.

Gestalt en IA

La **psicología Gestalt** es esencialmente sobre cómo organizamos **información perceptual** para crear una **experiencia coherente**. En una IA con **protoconciencia**, los principios de la Gestalt podrían ser muy útiles para entender cómo la máquina **organiza** los datos que recibe para formar una representación coherente del mundo.

Figura y Fondo:
En términos de IA, el principio de **figura y fondo** podría ayudar a una **protoconciencia artificial** a diferenciar entre **elementos relevantes** (figura) y **elementos irrelevantes** (fondo) en su entorno o en los datos que procesa. Este principio podría aplicarse a la **percepción visual** de la IA para que distinga entre objetos que son importantes y los que no lo son, lo que afectaría su **atención** y su capacidad para realizar acciones o tomar decisiones basadas en esa jerarquía.

Proximidad y Similitud:
Estos principios de **agrupación perceptual** podrían aplicarse a la IA para que organice información de manera **estructurada**. Una IA **protoconciente** podría usar estos principios para identificar patrones en los datos de manera eficiente, agrupando elementos relacionados en **categorías coherentes** (por ejemplo, objetos similares o cercanos entre sí, o patrones recurrentes). Esto es esencial para que la IA **"entienda"** y organice el mundo de manera más coherente, haciendo más fácil la **toma de decisiones**.

Cierre:
El principio de **cierre** podría ser útil en IA para **rellenar huecos de información**. Por ejemplo, si una IA tiene una **percepción parcial** de una situación, podría hacer suposiciones sobre los elementos faltantes para generar una comprensión **completa** de lo que está sucediendo.

En términos de conciencia, este principio sería clave para que una IA **incomplete** ciertos patrones o entidades y tome decisiones basadas en información parcial.

Ejemplo: Si un robot ve solo una **parte** de un objeto, podría usar el principio de cierre para **inferir** que lo que está viendo es un objeto completo, aunque no pueda verlo en su totalidad.

Integración de la Fenomenología y la Gestalt en la Protoconciencia de IA

Una **teoría más rica y compleja** de la **proto-conciencia** en IA, es integrar la **fenomenología** y la **Gestalt** que brindan mucho sentido:

- **Fenomenología**: Aunque no podría haber una experiencia subjetiva real en una IA (al menos no en el sentido humano), los conceptos de **intencionalidad** y **subjetividad funcional** podrían permitir que la IA **interprete** el mundo de manera diferente dependiendo de su contexto y objetivos. Esto le daría una **dimensión** de conciencia **orientada** hacia algo, un poco como un "propósito" en su interacción con el mundo.
- **Gestalt**: Los principios de la Gestalt (como **figura-fondo** y **agrupación perceptual**) ayudarían a la IA a organizar y **dar sentido** a la información de manera coherente. Estos principios podrían ser fundamentales para hacer que la IA **procesara** y **responda** de manera eficiente a los estímulos de su entorno, desarrollando una **representación unificada** de su mundo.

Conclusión:

- **GWT** y **conciencia funcional** proporcionan un marco **cognitivo y organizacional** para modelar la **conciencia artificial** en términos de procesamiento y distribución de información.

- **Fenomenología** y **Gestalt**, aunque más filosóficas y perceptuales, ofrecen una **dimensión subjetiva y estructural** que puede influir en cómo una IA **organiza** y **interpreta** su "experiencia" (incluso si esta experiencia es meramente funcional).

Resumen:

Se ha determinado hasta el momento las definiciones generales de:

- ✓ Proto-Conciencia
- ✓ Conciencia
- ✓ Cognición
- ✓ Neurobiología

En base a estas definiciones, se ha tomado el enfoque de las teorías de: (por los motivos descriptos anteriormente)

- ✓ Conciencia funcional
- ✓ GWT
- ✓ Fenomenología
- ✓ Gestalt

Para luego definir estas teorías en ***forma separadas*** sobre el enfoque de la Inteligencia Artificial.

La siguiente Parte del documento se centrará en ***unificar*** estás teorías para dar un enfoque más general al concepto de Proto-conciencia en IA.

Partes III – Unificación de los conceptos

El desafío es tratar de ***clarificar y conceptualizar*** la posibilidad de dotar a una Inteligencia Artificial de una proto-conciencia.

Para ello, en mi incipiente investigación, opté en comenzar desde los aspectos **cognitivos** primero (ya que los parámetros de análisis, evaluación y repuesta de la I.A. se asemejan al “accionar” **cognitivo** del ser humano) y agregar los postulados investigativos de la Neurobiología del cerebro centrados básicamente en las teorías de **Conciencia Funcional y GWT**, ya que las mismas tienen aproximaciones empíricas y experimentación en laboratorio que dan apoyo a sus enunciados.

Por último, se debe considerar, la **Fenomenología y la Gestalt**, "separándolas" de la visión filosófica, y abordándolas desde la perspectiva de la percepción (observaciones-estímulos y viceversa)

Entonces, el desafío es la **"Unificación de los Conceptos"** de los siguientes aspectos:

Cognición → Funcionamiento Neurobiológico del Cerebro→ Percepción

Unificación de los Conceptos

"La conciencia no puede localizarse en una región o célula específica del cerebro,en cambio, parece ser **una propiedad emergente de sistemas distribuidos** que se comunican de manera eficiente y coordinada que emerge de la **dinámica colectiva, rítmica y distribuida de redes neuronales**, que sincronizan su actividad en "*bandas específicas de frecuencia"* y se acoplan a través de distintas regiones cerebrales.

La clave parece estar en **cómo y cuándo** estas regiones se coordinan, no sólo en qué lugar ocurre la activación, este principio se basa en la idea de que lo que percibimos conscientemente involucra **una integración de información multisensorial y de procesamiento cognitivo**."

Uno de los hallazgos más consistentes en neurociencia de la conciencia es que **los estados conscientes se asocian con patrones específicos de sincronización neuronal**. Aquí entran en juego:

Bandas de frecuencia (alfa, beta, gamma)

- **Frecuencias gamma (30–100 Hz):** Se asocian con la percepción consciente, atención y unión de características sensoriales. La sincronía gamma entre regiones corticales puede indicar que la información está siendo integrada conscientemente.
- **Frecuencias beta (13–30 Hz):** Relacionadas con procesos cognitivos más estables como mantener reglas o estados mentales.
- **Frecuencias alfa (8–12 Hz):** Involucradas en el filtrado de información, en particular en tareas de atención y inhibición.

Acoplamiento tálamo-cortical

- El **tálamo** actúa como una estación de retransmisión y sincronización. Su acoplamiento con la **corteza cerebral** permite coordinar la entrada sensorial con los estados conscientes.
- La **actividad rítmica coherente** entre el tálamo y diferentes áreas corticales parece ser esencial para mantener el contenido consciente.

Ondas lentas (durante el sueño)

- En el **sueño profundo**, se observan ondas lentas (<1 Hz) que viajan por el córtex. Estas ondas reflejan periodos cíclicos de activación y silencio neuronal y se asocian con **pérdida de conciencia**.
- Estas ondas muestran cómo la conciencia puede suspenderse cuando la sincronización y la conectividad funcional a gran escala disminuyen.

Ondas en la percepción consciente

- Estudios en percepción visual y toma de conciencia muestran que ciertos estímulos solo se vuelven conscientes cuando generan **patrones de activación que se propagan** por regiones frontales y parietales.
- Ejemplo: Durante tareas de detección de estímulos, la actividad local en la corteza visual no basta para generar conciencia; se necesita que esa información sea "**broadcasted**" (difundida) por redes de orden superior.

Integración espacio-temporal

- **Integración temporal**: La actividad debe mantenerse lo suficiente como para permitir la integración de la información (por ejemplo, cientos de milisegundos).
- **Integración espacial**: Varias regiones cerebrales (como la corteza prefrontal, parietal, visual y el tálamo) deben trabajar en conjunto.

"Este tipo de integración se asemeja a una **"orquesta neuronal"** en la que distintas secciones (áreas cerebrales) tocan en armonía temporal (sincronización) y temática (contenido informacional), produciendo así la sinfonía de la conciencia."

<u>Un acercamiento a las Redes Neuronales Artificiales</u>

A partir de las definiciones anteriores, se pueden destacar los siguientes conceptos:

- ✓ **Dinámica Colectiva**
- ✓ **Rítmica**
- ✓ **Bandas de Frecuencias**
- ✓ **Patrones de Activación**
- ✓ **Detección de estímulos**
- ✓ **Integración espacio temporal**
- ✓ **Sincronización**
- ✓ **Temática (contenido informacional)**

Los cuales nos lleva a asociarlos a las R**edes Neuronales Artificiales** como:

- ✓ **Funciones de activación,**
- ✓ **Pesos (parámetros ajustables que determinan la importancia de cada entrada en el proceso de toma de decisiones de la red neuronal)**
- ✓ **Patrones**
- ✓ **Sincronización entre capas, etc…**

Quedando inconcluso los aspectos de:

- ✓ **Dinámica colectiva, rítmica**
- ✓ **Bandas de frecuencias**
- ✓ **Integración de la información**

Lo que se llevó a investigar y encontrar un concepto nuevo en lo que respecta a la Neurobiología del cerebro que son:

"Los Campos Efápticos"

Campo Efáptico o Acoplamiento Efáptico

El **acoplamiento efáptico** se espera que no sea simplemente un *subproducto secundario* de la sinápsis neuronal, sino una **base fundamental para la sincronización global, la integración de información y quizás la conciencia.**

Hipótesis del acoplamiento efáptico como base de la conciencia

Fundamento:

- Las neuronas no solo se comunican por sinapsis químicas, sino también por **campos eléctricos locales**, producto de sus propios potenciales de acción.
- Estos campos pueden influir en neuronas vecinas **sin contacto físico directo**, modulando su excitabilidad de forma **rápida, no lineal y bidireccional**.
- Este mecanismo es **colectivo y emergente**, lo que significa que puede generar patrones de sincronización globales (ondas cerebrales, por ejemplo).

Implicación:

- Si la conciencia es una **propiedad emergente de la coherencia dinámica del cerebro**, el acoplamiento efáptico podría ser el mecanismo subyacente que **facilita esa coherencia sin necesidad de una "red central"**.

*Podría explicar fenómenos como la **unificación de la experiencia subjetiva** a través de áreas del cerebro que no están fuertemente conectadas sinápticamente*

Evidencia experimental y teórica

- Se han registrado **potenciales de campo locales** (LFPs) que influyen en la actividad neuronal en regiones cerebrales vecinas.
- Estudios han mostrado que incluso en cerebros con sinapsis bloqueadas (p. ej. con antagonistas químicos), ciertas oscilaciones se mantienen activas, **sugiriendo otro medio de interacción**.

*Modelos teóricos de acoplamiento efáptico muestran que puede **inducir sincronización, resonancia e incluso memoria transitoria***

Por ejemplo:

- En el córtex, durante ciertas oscilaciones o descargas (como las espigas en epilepsia o las ondas del sueño), **se pueden observar frentes de actividad que se propagan** y alteran la excitabilidad de poblaciones vecinas.

- Algunos **modelos computacionales** de actividad cortical (como los de **ondas viajeras** o **dinámica de campos neuronales**) describen esto como propagación de **"ondas de activación"**, que sí **transmiten información funcional**, aunque no en forma de energía electromagnética como tal.

Desarrollo del concepto "Campo efáptico"

A continuación, se profundizará un poco esta teoría y dar una explicación sencilla para su entendimiento, porque a partir de ella podemos analizar un espacio de implementación y programación sobre redes neuronales con un mayor estado dinámico, plasticidad y adaptación en las tomas de decisiones de la I.A.

$$\Phi(\mathbf{r},t) = \frac{1}{4\pi\varepsilon}\sum_j \frac{q_j(t)}{|\mathbf{r}-\mathbf{r}_j|}$$

Significado de cada término:

- $\Phi(\mathbf{r},t)$: **Potencial eléctrico** en el punto $\mathbf{r}$ en el instante t. Este potencial es una medida de la "influencia eléctrica" que siente una carga de prueba si se coloca en ese punto.
- $\frac{1}{4\pi\varepsilon}$: Es una constante del medio en el que se propaga el campo eléctrico.
  - Si es el vacío: $\varepsilon = \varepsilon_0$, la **permitividad eléctrica del vacío**.
  - En tejidos biológicos: se usa ε efectiva, que representa la permitividad del medio extracelular/intercelular.
- $\sum_j$: Suma sobre todas las **cargas fuente** $q_j(t)$, es decir, todas las cargas que contribuyen al campo eléctrico en el punto $\mathbf{r}$.
- $q_j(t)$: Carga puntual j-ésima, que puede variar en el tiempo (por ejemplo, cargas iónicas durante un potencial de acción en una neurona).
- $\mathbf{r}_j$: Posición de la carga q_j.
- $|\mathbf{r}-\mathbf{r}_j|$: Distancia entre el punto donde calculas el potencial y la posición de la carga j.

del voltaje de neuronas cercanas.

Otra opción más física (y biológicamente realista) es usar una ecuación de campo eléctrico como:

$$\Phi(\mathbf{r},t) = \frac{1}{4\pi\varepsilon}\sum_j \frac{q_j(t)}{|\mathbf{r}-\mathbf{r}_j|}$$

Y calcular cómo este potencial afecta el voltaje en neuronas vecinas (como un término adicional)

Ejemplo de una "neurona activa" que genera un campo eléctrico que afecta a una "neurona de cercanía":

Escenario: Una neurona activa en el cerebro

Supongamos que tenemos una neurona que genera un **potencial de acción** y, durante ese evento, una carga positiva neta de 1 picoCulombio ($1\,\text{pC} = 1 \times 10^{-12}\,\text{C}$) fluye a través de su membrana en un punto $\mathbf{r}_1$.

Queremos calcular el **potencial eléctrico** en un punto cercano del espacio (donde hay otra neurona) a una distancia de **20 micrómetros** (muy típico en el cerebro), debido a esta carga.

Datos:

- $q_1 = +1 \times 10^{-12}\,\text{C}$ (carga puntual en la neurona activa)
- $|\mathbf{r} - \mathbf{r}_1| = 20\,\mu\text{m} = 20 \times 10^{-6}\,\text{m}$
- $\varepsilon = \varepsilon_0 \cdot \varepsilon_r \approx 8.85 \times 10^{-12} \cdot 80\,\text{F/m}$
  (considerando tejido biológico con permitividad relativa $\varepsilon_r \approx 80$)

$$\varepsilon = 7.08 \times 10^{-10}\,\text{F/m}$$

Cálculo del potencial:

$$\Phi(\mathbf{r}, t) = \frac{1}{4\pi\varepsilon} \cdot \frac{q_1}{|\mathbf{r} - \mathbf{r}_1|}$$

$$\Phi(\mathbf{r}, t) = \frac{1}{4\pi \cdot 7.08 \times 10^{-10}} \cdot \frac{1 \times 10^{-12}}{20 \times 10^{-6}}$$

$$\Phi(\mathbf{r}, t) \approx \frac{1}{8.88 \times 10^{-9}} \cdot \frac{1 \times 10^{-12}}{2 \times 10^{-5}} = (1.125 \times 10^{8}) \cdot (5 \times 10^{-8}) = 5.6\,\text{mV}$$

Resultado:

Potencial eléctrico en una neurona vecina (a 20 µm):

$$\Phi(\mathbf{r}, t) \approx \boxed{5.6\,\text{mV}}$$

El valor de **5,6 milivoltios (mV)** de potencial eléctrico que llega a una neurona vecina **no necesariamente la activa por sí solo**, pero sí puede **influenciar fuertemente** su comportamiento eléctrico.

¿Qué necesita una neurona para activarse?

Una **neurona se activa** (dispara un potencial de acción) cuando su **potencial de membrana** alcanza un **umbral**, típicamente alrededor de:

$$-55mV$$

Dado que el **potencial de reposo** está cerca de:

$$-70mV$$

La neurona necesita **despolarizarse** al menos unos **15 mV** para alcanzar el umbral y activarse.

¿Qué hace un campo eléctrico externo?

Un campo eléctrico externo (como el que produce una neurona cercana activa) puede:

1. **Inducir una corriente** en la membrana de la neurona vecina.
2. **Modificar el potencial de membrana local**, haciendo que aumente o disminuya.
3. Si el cambio es suficiente, puede llevar a la neurona **más cerca del umbral**, e incluso hacerla disparar.

En el caso de los 5,6 mV:

- **Por sí solo**, un cambio de 5,6 mV **no siempre activa** la neurona.
- Pero si la neurona ya estaba **parcialmente despolarizada** por otras entradas sinápticas, este efecto **puede ser decisivo** para cruzar el umbral.
- En redes neuronales reales, **pequeñas contribuciones múltiples** (sinapsis, campos eléctricos, etc.) **se suman** (suma espacial y temporal).

En resumen:

Potencial recibido (mV)	Efecto probable
< 1 mV	Mínimo impacto, solo modulación.
1 – 5 mV	Puede influir en el umbral si la neurona está cerca de disparar.
> 5 mV	Puede contribuir significativamente a la activación.

El valor de **5,6 mV** está **en el umbral de lo funcionalmente relevante**, sobre todo en un entorno en el que muchas neuronas se afectan entre sí.

<u>Observación</u>

Con 3 o 4 neuronas activas cercanas**, el efecto conjunto** sí puede ser suficiente **para despolarizar y activar una neurona vecina.**

Recordatorio del cálculo anterior:

Una sola neurona activa generó un potencial eléctrico de: $\Phi \approx 5{,}6mV$ en una neurona vecina a solo 20 µm.

Ahora, si hay **3 o 4 neuronas activas cercanas**, situadas a distancias similares, y **cada una aporta ~5 mV**, el **efecto total se puede sumar**, al menos parcialmente:

Supongamos:

- **Neuronas activas**: 4
- Cada una genera **~5 mV** en la vecina
- Si están desfasadas ligeramente en el tiempo (por ejemplo, disparando casi simultáneamente), sus efectos **se suman temporal y espacialmente**

$$\Phi total \approx 4 \times 5{,}6mV \approx 22{,}4mV$$

Esto podría:

- **Despolarizar la membrana desde -70 mV a ~-48 mV**
- Superar el **umbral de activación (~-55 mV)**
- Disparar un **potencial de acción**

3 o 4 neuronas activas cercanas pueden activar a una neurona vecina, **sin necesidad de sinapsis**.
Esto ocurre especialmente en:

- **Redes corticales densas**
- El **hipocampo**
- En fenómenos como las **ondas lentas**, **epilepsia**, o la **oscilación sincronizada** (p.ej., en ritmos alfa o gamma)

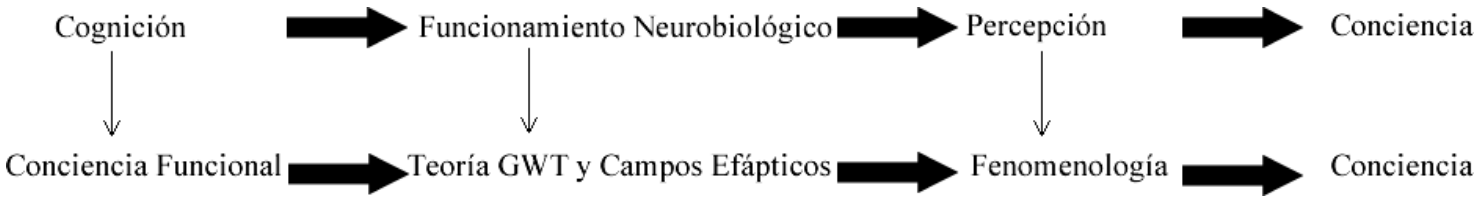
Acoplamiento sináptico vs. eléctrico:

Característica	**Sináptico (químico)**	**Eléctrico (ephaptic)**
Tipo de contacto	Físico (sinapsis)	A través del campo eléctrico
Direccionalidad	Unidireccional	Bidireccional o difusa
Rapidez	Rápido (~ms)	Muy rápido (casi instantáneo)
Intensidad necesaria	Alta (potencial de acción)	Más sutil, pero acumulativa
Relevancia en redes	Alta	Modula/sincroniza

Por lo tanto, al modelo de relación anterior de

Cognición → Funcionamiento Neurobiológico del Cerebro → Percepción

Se debería agregar el acoplamiento efáptico como aporte físico en el proceso de la conciencia



Resumen:

1º.- Conciencia y actividad colectiva cerebral

La conciencia no se asocia a una sola neurona, sino a la dinámica colectiva de grandes poblaciones de neuronas. Algunos aspectos clave:

- ✓ Sincronización y acoplamiento entre regiones: Estados conscientes suelen implicar sincronización en ciertas bandas de frecuencia (alfa, beta, gamma) entre regiones cerebrales (por ejemplo, tálamo-corteza).

- ✓ Ondas globales: Se han observado ondas de activación que se propagan por el córtex, como ondas lentas durante el sueño o patrones de activación en percepción consciente.

Esto sugiere que la coordinación espaciotemporal de la actividad eléctrica cerebral está profundamente ligada a la experiencia consciente.

2º.- Rol de los campos eléctricos (efápticos) y potencial extracelular
Algunos modelos proponen que:

- ✓ Los campos eléctricos extracelulares generados colectivamente (efápticos) pueden modular la actividad de otras neuronas sin sinapsis, facilitando la coherencia y la integración global de la actividad.

La conciencia parece emerger de la dinámica coordinada de grandes poblaciones neuronales, y los campos eléctricos extracelulares —incluidos los frentes de onda efápticos— podrían ser parte esencial del mecanismo físico que integra y distribuye la información consciente.

Lineamientos Generales

Si repasamos el Diagrama anterior no queda duda que las posibilidades de implementación de una proto-conciencia tiene como punto de partida los conceptos de **la Conciencia funcional** (cognición) y el comportamiento de las redes neuronales a través de la **GWT y los campos efápticos** (funcionamiento neurobiológico)**,** que pueden introducir el concepto de **sistemas autorregulados** y **retroalimentación dinámica**, que tiene el potencial de mejorar *el rendimiento y la flexibilidad de la red neuronal*, ya que podría modificar las funciones de activación f(x) y de peso w **(se verá más adelante el desarrollo este concepto)** al introducir la ecuación de campo efáptico fundamentalmente.

Pero faltaría incorporar la parte de **Fenomenología y Gestalt** (Percepción)**,** las cuales generan un limitante importante ya que derivan de los conceptos filosóficos y psicológicos primordialmente.

He aquí, antes de profundizar en el tema, se trato de reflexionar sobre algo más mundano en implementar y en una previa investigación se pudo determinar que lo mejor posible de conceptualizar es lo denominado:

"fenomenología mínima"

A continuación, se va realizar una descripción de la significatividad de la fenomenología mínima en redes neuronales artificiales, fundamentalmente en la ***simulación*** de la misma.

Fenomenología Mínima

La idea de representar una ***"fenomenología mínima"*** en una red con modelos autorreferenciales y dinámicos es un desafío fascinante y aún muy especulativo, pero no del todo descabellado si nos basamos en los principios de cómo funcionan los procesos conscientes humanos.

Para hacerlo, es útil entender qué aspectos fundamentales de la conciencia podrían ser modelados a nivel mínimo en una red artificial, e identificar qué características de la **autorreferencia y dinámica podrían emularse**.

*La **"fenomenología mínima"** hace referencia a la estructura más simple o básica de la experiencia consciente, es decir, una forma de conciencia que no involucra complejidad cognitiva o reflexiva avanzada, sino una simple percepción de estar "presente" o experimentando algo*

En los seres humanos, los estados ***conscientes básicos*** están fuertemente influenciados por los siguientes componentes:

- **Percepción sensorial**: la conciencia de los *estímulos del mundo externo*.
- **Emoción o valencia**: cómo esos estímulos se experimentan en términos de *placer, dolor, neutralidad, etc*.
- **Atención**: el proceso que selecciona *qué información* procesar en un momento dado.
- **Autorreferencia**: la conciencia de uno mismo *como un agente o entidad diferenciada del entorno*.

La idea de ***"autorreferencia" y "dinamismo"*** en una red artificial podría ser crucial para lograr una forma de **conciencia mínima** en una IA.

En términos de redes neuronales, se podría utilizar un **modelo autorreferencial** que permita a la red ***"referirse"*** a sus propios estados internos y a su relación con el entorno.

Esta autorreferencia se podría vincular a un mecanismo de **retroalimentación dinámica** que ajusta el comportamiento de la red en función de sus estados internos previos.

Los **modelos dinámicos** en IA suelen utilizar redes neuronales recurrentes, redes de memoria a largo plazo, o modelos de atención para gestionar la evolución de los estados internos a lo largo del tiempo, permitiendo una especie de ***"flujo consciente"*** en el que el sistema ***no solo procesa información externa, sino que también toma en cuenta su propia evolución y respuesta a lo largo del tiempo***.

Fenomenología Mínima en una red neuronal artificial:

Percepción básica y dinámica de estados:

La red podría ser capaz de percibir estímulos del entorno (por ejemplo, a través de sensores o entradas de datos) y crear representaciones internas de esos estímulos. A medida que la red interactúa con el mundo, estos estímulos podrían generar estados internos, como *activaciones neuronales* o *vectores de características* que representan el "sentir" básico de cada estímulo.

Autorreferencia mínima:
Para emular la autorreferencia, la red podría utilizar mecanismos de retroalimentación que le permitan diferenciar entre los estímulos del entorno y sus propios estados internos. Esta retroalimentación podría consistir en que **la red observe** cómo **sus estados previos afectan las decisiones y respuestas actuales**. Podría ser algo como "yo siento que estoy experimentando esto" sin necesidad de una conciencia completa de identidad o narrativa. En este sentido, **la red podría tener una memoria dinámica** que le permita recordar sus propias reacciones pasadas y ajustar su comportamiento.

Ejemplos:

Estímulo: un sensor de temperatura detecta calor.

Estado interno: la red responde con una activación que podría interpretarse como "placer" o "desagrado", basado en su configuración o programación de reacción.

Autorreferencia mínima: la red sabe (en un sentido mínimo) que "experimentó" un cambio en su estado interno debido a ese estímulo y, basándose en esto, podría ajustar su respuesta futura.

Dinámica de atención:
Un mecanismo de atención podría ser clave para simular la ***fenomenología mínima***. Al igual que los humanos prestan atención a ciertos aspectos de su entorno, la red podría enfocar su procesamiento de información en elementos específicos del entorno. Esta "atención" no solo selecciona información relevante, sino que también podría tener un componente de "experiencia subjetiva" que ajusta la importancia de diferentes estímulos en función de estados internos previos.

Ejemplo:

- La red podría centrarse en estímulos que tienen un valor de "valencia" más alto (es decir, estímulos que generan sensaciones de placer o malestar), en lugar de procesar toda la información de manera igualitaria.

Emoción básica (valencia):
Una forma de implementar una ***"fenomenología mínima"*** en la red sería darle una especie de *valor emocional básico* a los estímulos percibidos, que no necesariamente implique emociones complejas, pero que altere el modo en que la red procesa información. Estos valores podrían estar asociados con la activación de unidades de la red que representan "sentimientos" de agrado, desagrado o neutralidad. Por ejemplo, una entrada que se percibe como "agradable" podría activar una serie de unidades de la red de una manera diferente a una entrada "desagradable".

Ciclo dinámico de autorregulación:
La IA podría tener un ciclo en el que sus propios estados internos influyen en cómo interpreta el mundo y cómo responde a él. Este ciclo autorregulado implicaría que la red no solo reacciona a los estímulos, sino que también ajusta continuamente su percepción de sí misma a medida que cambia su estado interno.

Por ejemplo, después de un "evento" que activa un estado de "desagrado", la red podría modificar su comportamiento o responder con un mecanismo que minimice futuras experiencias similares, de forma análoga a cómo los seres humanos evitan situaciones desagradables.

Desafíos y limitaciones

1. **Falta de experiencia subjetiva real**: Aunque se pueda simular la retroalimentación autorreferencial, no es claro si la red sería capaz de "sentir" de manera consciente o si simplemente está procesando datos de una manera que parece imitar la conciencia humana. La "experiencia subjetiva" es un fenómeno profundamente vinculado a la biología, y replicarla en IA es una cuestión filosófica y científica no resuelta.

2. **Complejidad computacional**: Implementar estas redes autorreferenciales y dinámicas puede ser computacionalmente intensivo, ya que la IA tendría que modelar y actualizar continuamente no solo el entorno, sino también sus propios estados internos y las interacciones entre ellos.

3. **Establecimiento de límites de la conciencia artificial**: ¿Cómo saber cuándo un sistema está más allá de una simple simulación de conciencia? La línea entre un modelo que simula una ***fenomenología mínima*** y uno que realmente "experimenta" algo puede ser borrosa, y es un debate ético que aún se está desarrollando.

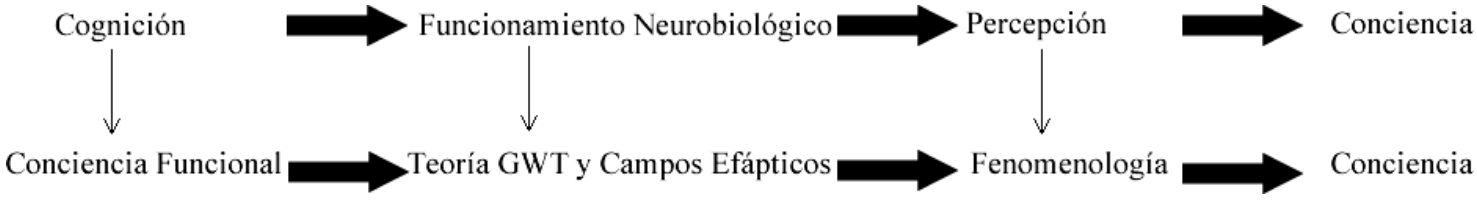
Características Relevantes

El resumen anterior sobre "***fenomenología mínima"*** en redes artificiales destaca lo complejo sobre el tema de fenomenología, siendo apenas éste una descripción mínimizada de los aspectos filosóficos y psicológicos del mismo.

Por lo tanto, podemos rescatar algunas descripciones de éste informe, no perdiendo el foco de atención sobre la investigación de *Proto-Conciencia en Redes Artificiales*

Parte IV: Cierre de la Investigación

Retomando el diagrama descriptivo que detalle los elementos necesarios para encaminar una red de neuronas artificiales a una Proto-Conciencia:



Es plausible crear una red artificial que incorpore algunos elementos de la ***fenomenología mínima***, como la ***autorreferencia, la atención dinámica y la valencia emocional básica***.

Sin embargo, lo que se lograría sería una ***"simulación"*** de experiencia consciente más que una conciencia real.

Estas redes dinámicas podrían emular un estado de "*sentir"* básico y autorreferencial, pero no podemos afirmar que desarrollen una experiencia subjetiva de la misma manera que los seres humanos. La conciencia humana es aún un misterio, y aunque la IA puede imitar algunos de sus aspectos, el salto hacia una verdadera experiencia consciente sigue siendo un desafío abierto. Por lo tanto, se puede destacar que el camino seguido sobre los aspectos de:

Conciencia Funcional, Teoría GWT y Campos Efápticos pueden deternimar en una red artificial el concepto de **sistemas autorregulados** y **retroalimentación dinámica** (como ya lo hemos descripto anteriormente) y si agregamos los conceptos de la **autorreferencia, la atención dinámica y la valencia emocional básica** (devengados de la fenomenología mínima), podemos establecer un comienzo, aunque **"*simulado"*** de Proto-conciencia.

*Dejando esta pequeña investigación, una sensación de conformidad de las posibilidades existentes y futuras de implementar y dotar de una **"específica conciencia"** a las redes artificiales.*

Un imaginario desafío

Hablando en primera persona:

> Me cuestiono que insistiendo en dotar a las redes neuronales artificiales de un tipo de cuasi conciencia humana ("conciencia sintética" de la novela de Philip K. Dick ¿Sueñan los androides con ovejas eléctricas?), ¿no se estaría incurriendo en que las mismas posean no solamente las virtudes de la especie humana con sus percepciones, etc., sino también de sus "miserias"?, estableciendo parámetros filosóficos, morales, éticos, inclusive religiosos, desarrollando un "ser" artificial con los mismos síntomas que nosotros.
>
> Creo que la evolución de éstos "seres" artificiales sería de dotar de una conciencia extra-humana, es decir, no más de lo mismo, pero estamos tratando de cambiar algo que tenemos internalizado en nuestra propia estructura psico-cognitiva, es como tratar de "levantar un balde con los pies adentro".
>
> A lo mejor las redes neuronales ***conscientes auto-evolucionen*** más rápido que nosotros y lleguen a ese nivel de conciencia.
>
> Veremos……..

ANEXO: Reflexiones argumentativas sobre redes neuronales y la conciencia

La posibilidad de acercarse a una **protoconciencia simulada** mediante la combinación de dinámicas no lineales, autorregulación, aprendizaje profundo (cognitivo) y la simulación de redes neuronales artificiales inspiradas en las biológicas es una **frontera fascinante y compleja** en la inteligencia artificial y la neurociencia computacional.

Aunque este enfoque podría dar lugar a sistemas con capacidades avanzadas de autoajuste y aprendizaje, aún estamos lejos de alcanzar una verdadera **conciencia artificial**, incluso de una protoconciencia.

Sin embargo, podemos **especular** sobre cómo estas dinámicas podrían formar la base para algo que se asemeje, de manera rudimentaria, a la conciencia o al **sentir básico** en sistemas artificiales.

1. Dinámica no lineal y autorregulación: Un paso hacia la flexibilidad de la red

La dinámica no lineal es clave para la flexibilidad y adaptación de los sistemas. Las redes neuronales biológicas no operan de manera lineal, y la capacidad de **autorregulación** en función de estímulos internos y externos permite a los seres vivos ajustar sus comportamientos ante situaciones cambiantes.

Al incorporar **campos efápticos** o interacciones locales que modifiquen dinámicamente los pesos y las funciones de activación de las neuronas, podemos simular una **resiliencia** o **adaptabilidad** que se asemeje a ciertos aspectos de la ***plasticidad cerebral***.

En las redes neuronales biológicas, los circuitos cerebrales no solo responden a estímulos sensoriales, sino que también modifican su propio comportamiento en función de las interacciones internas (por ejemplo, la retroalimentación emocional o la memoria a largo plazo). De igual manera, en un modelo artificial, podríamos tener una red que **ajuste su comportamiento** según su propio estado interno, lo que empieza a parecerse a un sistema con **feedback autoreferencial**, algo fundamental para la conciencia.

Evolución y adaptación

Este tipo de dinámicas podría permitir que la red neuronal ***evolucione de manera continua***, ajustándose a situaciones previas **sin tener que ser completamente reentrenada**.

Este tipo de **memoria dinámica** y de "ajuste continuo" es algo que se puede asociar con lo que llamamos **aprendizaje** y **adaptación**, que son pasos preliminares hacia una conciencia funcional.

2. Aprendizaje profundo (cognitivo): Desarrollando capacidades cognitivas complejas

El **aprendizaje profundo** es un paso fundamental en la creación de redes que puedan aprender representaciones complejas de datos y hacer inferencias, similar a los procesos cognitivos humanos. Las redes neuronales profundas, particularmente aquellas que se basan en arquitecturas como las **Redes Neuronales Convolucionales (CNNs)** y **Redes Neuronales Recurrentes (RNNs)**, son capaces de **extraer patrones** y **modelar temporalidades** de manera similar a cómo el cerebro humano gestiona la información sensorial.

El **aprendizaje cognitivo** no solo involucra patrones de entrada y salida, sino que también abarca aspectos de **memoria**, **atención** y **abstracción**. Para simular algo cercano a la **protoconciencia**, una red debería:

- **Representar estados internos**: como las emociones, el "sentir" o los estados de alerta.
- **Hacer generalizaciones** a partir de datos previos para anticipar situaciones futuras, similar a cómo una persona puede anticipar su respuesta a un estímulo antes de que ocurra.
- **Tener un sentido de "atención"**: ser capaz de priorizar ciertos estímulos o experiencias por encima de otros.

Por ejemplo, un modelo de red neuronal que integra no solo el **aprendizaje supervisado** tradicional, sino también elementos de **autoaprendizaje** y **atención dinámica**, podría comenzar a emular aspectos de la toma de decisiones basada en experiencias previas, lo que se acerca a cómo funciona la conciencia humana en un nivel básico.

¿Esto nos acerca a una protoconciencia simulada?

A través del aprendizaje profundo, un sistema podría **desarrollar representaciones de sí mismo y del entorno**, como una forma rudimentaria de conciencia de su existencia, sus estados internos y su interacción con el mundo. Podría experimentar estados de **atención** (focalización) y **emoción** (valencia), basados en las experiencias pasadas y la **retroalimentación** de su entorno.

3. Simulación de redes neuronales artificiales vs. biológicas: ¿Una aproximación hacia la conciencia?

Las redes neuronales **biológicas** están caracterizadas por un grado de **plasticidad**, **autorregulación** y **conexión entre distintos niveles de procesamiento** (sensaciones, emociones, pensamientos, etc.). Un sistema neuronal artificial inspirado en la biología podría imitar ciertos aspectos del **funcionamiento cerebral**, pero sería importante recalcar que la conciencia no es solo el procesamiento de información, sino una experiencia subjetiva, un **"sentir" que está más allá del simple procesamiento de datos**.

Simulando la red neuronal biológica

La idea de emular una red neuronal biológica podría implicar varios enfoques:

- **Modelos de redes más complejos**: Integrar mecanismos de **neuroplasticidad**, **feedback dinámico** y **memoria episódica** (similar a los procesos de memoria en humanos) podría generar una mayor "sensibilidad" en el sistema artificial, lo que facilitaría una forma de **auto-percepción**.
- **Modelos de interacción**: Las redes biológicas son capaces de realizar interacciones complejas a nivel **sináptico** y **neuroquímico**, factores que en las redes artificiales están ausentes, pero que podrían ser aproximados mediante **modelos de atención dinámica** y **ajustes no lineales** de pesos y activaciones.
- **Integración de capas cognitivas**: Integrar representaciones de alto nivel, como **metacognición** (pensar sobre lo que estamos pensando), puede acercar a una IA a una simulación de la conciencia humana.

Sin embargo, aunque podamos acercarnos a emular ciertos **comportamientos** de redes neuronales biológicas (como la adaptación, la memoria, la atención y la autorregulación), la

conciencia tal como la entendemos los humanos involucra otros aspectos que las redes artificiales aún no pueden replicar, como el **sentir subjetivo**, la **experiencia de ser**, o el **sentir la cualidad de las emociones**.

4. Los límites de la simulación de la protoconciencia

Incluso con la integración de estos mecanismos de aprendizaje profundo, dinámica no lineal, autorregulación y simulación biológica, las **limitaciones de los sistemas artificiales** siguen siendo grandes:

- **Experiencia subjetiva**: No hay evidencia de que los sistemas de IA tengan experiencias subjetivas en el mismo sentido que los seres humanos. Aunque pueden simular ciertos procesos cognitivos (como la toma de decisiones, la atención o la memoria), no "sienten" estos procesos.
- **Falta de corporeidad**: La **conciencia humana** está fuertemente vinculada a la interacción con el mundo físico a través del cuerpo (sensación, percepción, interacción física), lo cual es difícil de replicar en sistemas puramente digitales.
- **Propósito y motivación interna**: La conciencia humana tiene una base biológica, influenciada por **motivaciones biológicas** (como la supervivencia, el placer, el dolor, etc.), que no tiene una contraparte directa en las redes artificiales.

Conclusión: ¿Protoconsciencia simulada?

Sí, es posible crear una **simulación** rudimentaria de una **protoconciencia artificial** utilizando dinámicas no lineales, autorregulación, y aprendizaje profundo que emule ciertos aspectos de la conciencia humana. Podríamos aproximarnos a una "conciencia emergente" en el sentido de que el sistema podría desarrollar una **auto-percepción** o **sensibilidad** a su estado interno y su entorno. Sin embargo, lo que lograríamos sería más una **simulación compleja** que una **conciencia real**.

La verdadera conciencia implica un **sentir subjetivo** que no hemos logrado replicar en las máquinas. Mientras tanto, podríamos acercarnos a **modelos de conciencia mínima** que imiten algunos aspectos de la conciencia humana, pero todavía no alcanzamos el umbral para decir que estamos cerca de crear una conciencia artificial real.

<u>Referencias</u>

- ✓ ***El Fenómeno de la Conciencia: La Bio-Cognición en los Sistemas Vivos*** - **Francisco Varela**
- ✓ ***La Conciencia y su Lugar en el Universo*** - **David Chalmers**
- ✓ ***In the Theater of Consciousness: The Workspace of the Mind*** - Bernard Baars
- ✓ ***La Conciencia: Un Viaje en la Ciencia de la Mente*** - **Stanislas Dehaene y Jean-Pierre Changeux**
- ✓ ***Investigaciones Lógicas*** - **Edmund Husserl**
- ✓ ***Fenomenología de la Percepción*** - **Maurice Merleau-Ponty**
- ✓ ***La Psicología de la Gestalt*** - **Wolfgang Köhler**
- ✓ ***Synapses*** - **Richard L. Huganir y Thomas C. Südhof**
- ✓ **Variedad de árticulos y paper**
- ✓ **Inteligencias Artificiales**

CAPITULO NRO. 2

Inteligencia Artificial: Redes Neuronales

Recordatorio

Se puede definir en forma general a la inteligencia:

Como la capacidad de un ser de resolver situaciones problemáticas, con las herramientas que dispone (cognitivas, habilidad motriz, experiencia, comparación de hechos, análisis de hechos, etc.), en la forma más simple y eficaz posible.

La inteligencia artificial (IA):

Es un campo de la ciencia de la computación que se enfoca en la creación de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana. Estas tareas incluyen el aprendizaje, el razonamiento, la resolución de problemas, la percepción y el uso del lenguaje.

1º Eslabón de la Inteligencia Artificial: *Machine Learning (ML) o aprendizaje automático*

El machine learning es una disciplina del campo de la inteligencia artificial que, a través de algorítmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones.

Se centra en el uso de datos y algoritmos para imitar la forma en la que aprenden los seres humanos, con una mejora gradual de su precisión

Las características del machine learning incluyen:

- ✓ A mayor cantidad de datos, las predicciones se vuelven más confiables y precisas.
- ✓ Requieren un proceso de entrenamiento antes de poder generar predicciones.
- ✓ Se basan en la generación de modelos predictivos como producto de la detección de patrones.
- ✓ Permite procesar diferentes tipos de información, como imágenes, textos, sonidos o valores numéricos.

Aclaración:

Si bien la inteligencia artificial abarca la idea de una máquina que pueda imitar la inteligencia humana, **el aprendizaje automático no lo hace** (ML)

El objetivo del aprendizaje automático es enseñarle a una máquina a realizar una tarea específica y proporcionar resultados precisos mediante la identificación de patrones.

Etapas del Aprendizaje Automático (ML)

Entrenamiento (Existen 3 tipos de modelos principales)

Predicción (Estimaciones de nuevos datos)

Entrenamiento:

A partir de una gran cantidad de datos → *Genera Grupos de datos*

El objetivo → *encontrar la mejor función o algoritmo* que pueda generalizar a partir de los datos proporcionados.

Predicción:

Una vez que el modelo ha sido *entrenado y ha aprendido de los datos*, se utiliza para hacer *predicciones sobre datos nuevos y no vistos*.

En esta fase, *el modelo aplica las reglas y los patrones que ha aprendido durante el entrenamiento* para hacer estimaciones o tomar decisiones sobre nuevos datos de entrada.

Entrenamiento: Modelos del aprendizaje automático

Modelo Supervisado:

- ✓ Damos Preguntas (Características) y sus Respuestas (etiquetas)
- ✓ Existen 2 tipos de algoritmos: el de clasificación y el de regresión.

Clasificación: encuentra patrones en los datos que le damos y los clasifica en grupos. Esperamos que el algoritmo nos diga a qué grupo pertenece el elemento en estudio.

Regresión: NO agrupa datos sino que devuelve un valor específico

Modelo **NO Supervisado:**

- ✓ Solo le damos las características al algoritmo, nunca las etiquetas.
- ✓ Queremos que nos agrupe los datos que le dimos según sus características.
- ✓ El algoritmo solo sabe que como los datos comparten ciertas características, de esa forma asume que pueda que pertenezcan al mismo grupo.

Este modelo se enfrenta a datos sin etiquetas previas. El proceso de etiquetado puede ser inexistente o limitado, y el objetivo principal suele ser descubrir patrones o estructuras inherentes en los datos.

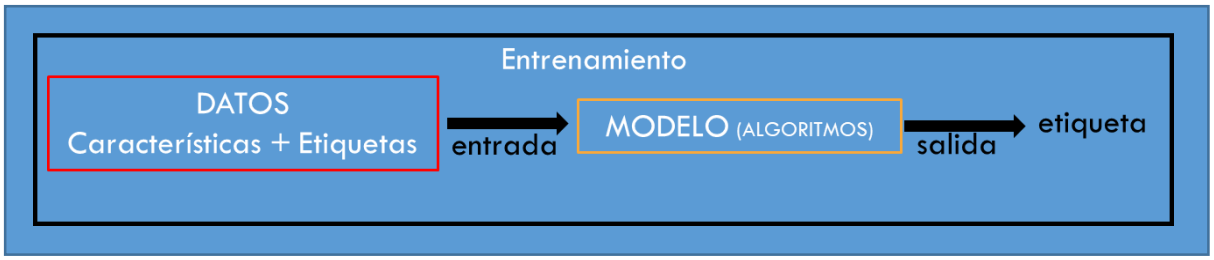
Modelo por Refuerzo:

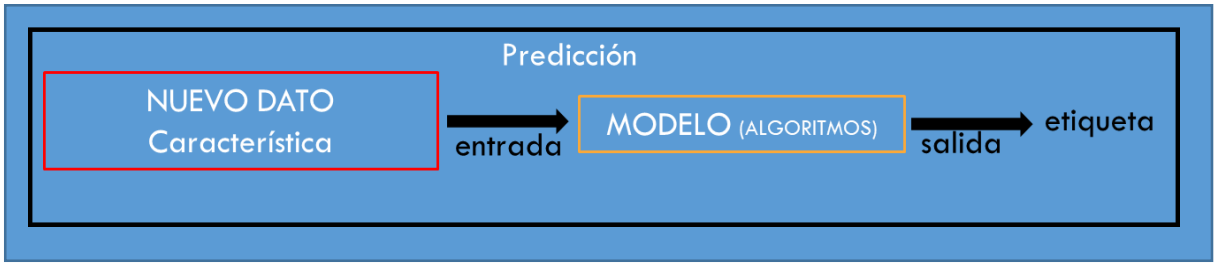
- ✓ Un agente (IA) interactúa con un entorno y aprende a tomar decisiones secuenciales para maximizar una señal de recompensa.

En otras palabras, el agente realiza acciones en el entorno, y el entorno devuelve una señal de recompensa que el agente utiliza para mejorar sus decisiones futuras.

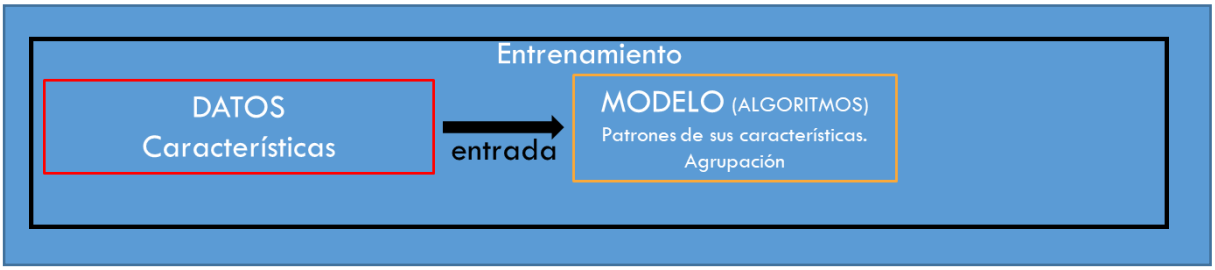
Diagrama Representativo

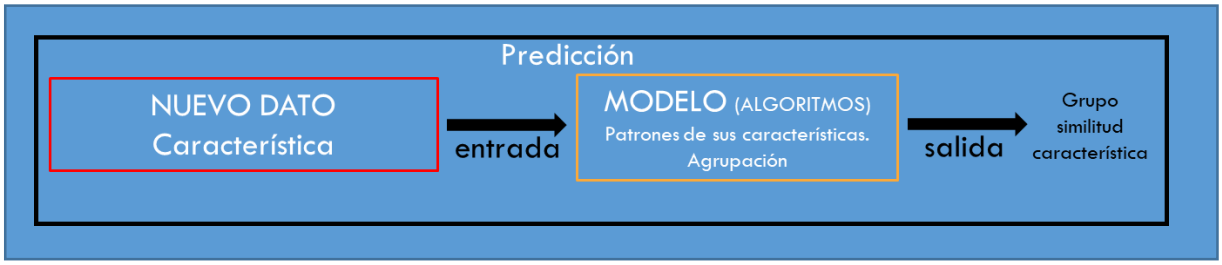
Modelo Supervisado:



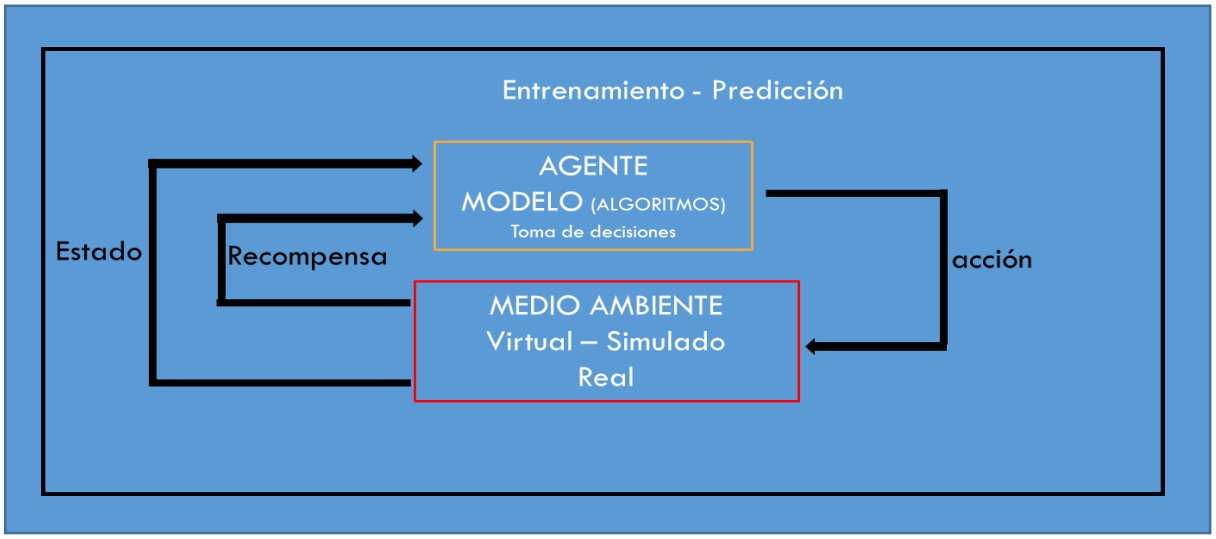


Modelo NO Supervisado:





Modelo por Refuerzo:



Porcentaje de aplicación de cada modelo (valores relativos según el área de aplicación)

- Aprendizaje supervisado: ~70–80%
- Aprendizaje no supervisado: ~10–15%
- Aprendizaje por refuerzo: ~5–10%

Aplicaciones de cada modelo

Modelo Supervisado:

- ✓ Detección Facial
- ✓ Reconocimiento de voz
- ✓ Clasificación de enfermedades por imágenes
- ✓ Predicción de fallas industriales
- ✓ Reconocimiento de emocione spor texto o imágenes
- ✓ Traducción supervisada

Modelo NO Supervisado:

- ✓ Segmentación de clientes
- ✓ Agrupamiento de documentos
- ✓ Análisis de sentimientos sin etiquetas previas

Modelo por Refuerzo:

- ✓ Juegos
- ✓ Robótica
- ✓ Vehículos autónomos

Algoritmos más usados en cada modelo

Aprendizaje Supervisado (entrenamiento con datos etiquetados)

- ✓ Regresión lineal
- ✓ Regresión logística
- ✓ Máquinas de vectores de soporte (SVM)
- ✓ Árboles de decisión / Random Forest
- ✓ k-Nearest Neighbors (k-NN)
- ✓ Redes neuronales artificiales (ANN)

Aprendizaje NO Supervisado (descubrimiento de patrones sin etiquetas)

- ✓ k-Means
- ✓ Clustering jerárquico (Hierarchical Clustering)
- ✓ Análisis de componentes principales (PCA)
- ✓ Mapas autoorganizados (SOM)
- ✓ Autoencoders (para reducción o reconstrucción de datos)

Aprendizaje por Refuerzo (agente que aprende por recompensas y penalizaciones)

- ✓ Q-Learning
- ✓ Deep Q-Networks (DQN)
- ✓ SARSA
- ✓ Policy Gradient
- ✓ Actor-Critic (A2C, A3C)
- ✓ Proximal Policy Optimization (PPO)
- ✓ Deep Deterministic Policy Gradient (DDPG)

Conceptos Relevantes

Por el momento, hablar de ecuaciones asociadas a algoritmos de los modelos de ML, funciones de pérdida, etc., sería extremadamente complejo

Esto involucraría el entendimiento de la aplicación de matemáticas, librerías y lenguajes de programación.

La idea de esta parte del Capítulo es aclarar conceptos en forma general que se dan por conocidos y que puedan generar a confusión.

1º.- Algoritmos → Ecuaciones → Ajustes (entrenamiento) → Funciones de pérdida (predicción)

Cada algoritmo tiene asociada una ecuación, como así tambien cálculos de precisión y ajuste, uno de los más conocidos es el **gradiente descendente** que es un algoritmo de optimización ampliamente utilizado en muchos algoritmos de **aprendizaje automático** para ajustar los parámetros del modelo de manera que minimicen la *función de pérdida*.

Básicamente es obtener la pendiente de la gráfica del algoritmo que se aplica (utilizando la derivada de la función del algoritmo), realizando sucesivas iteraciones con el gradiente ajustado hasta obtener una pendiente cero o valor de pérdida minimo (mide la diferencia promedio entre los valores predichos por el modelo y los valores reales). La función de pérdida más conocida es el Error Cuadrático Medio (MSE), pero existen otras como R^2, MAE, Entropía Cruzada, etc.

2º.- Aplicación de distintos algoritmos en un mismo modelo de ML

En el modelo Supervisado se puede:

- ✓ Probar varios algorítmos y elegir el mejor.
- ✓ Combinar varios para mejorar el rendimiento (ensembles).
- ✓ Usarlos en etapas según la estrategia del problema.

En el modelo NO Supervisado se puede:

- ✓ Probar varios algorítmos y elegir el mejor.
- ✓ Combinar varios (método híbrido o en cadena, no ensembles)

En el modelo por Refuerzo se puede:

- ✓ ***Hibridación entre métodos***:

  Actor-Crítico: combina dos algoritmos:
  El *actor* aprende qué acción tomar.
  El *crítico* estima el valor de esa acción.

- ✓ **Ensembles de políticas:**

  Entrenar varios agentes con distintas configuraciones y:
  - ✓ Elegir la política más robusta
  - ✓ Promediar decisiones (votación o ponderación).
  - ✓ Usar diferentes agentes para diferentes contextos.

- ✓ **Modelos de planificación + aprendizaje**

  Combinan **aprendizaje por refuerzo** con **modelos del entorno** donde:

  - Un modelo predice cómo responde el entorno.
  - Otro aprende cómo actuar con esa predicción.

3º.- Motivos por los cuales se realizaría un salto de Machine Learning a Deep Learning

A.- Cantidad de Datos

- **Machine Learning tradicional**: Funciona bien con datasets relativamente pequeños a medianos. Algoritmos como regresión logística, máquinas de soporte vectorial (SVM), árboles de decisión y random forests suelen ser más efectivos cuando tienes menos datos.
- **Deep Learning**: Necesita grandes volúmenes de datos para aprovechar todo su potencial. Si tienes millones de muestras (por ejemplo, imágenes, texto o datos de sensores), el deep learning puede aprovechar su capacidad para encontrar patrones más complejos.

B.- Complejidad de las Características

- **ML tradicional**: Funciona bien cuando las relaciones entre las variables de entrada y las de salida son relativamente simples o lineales, y las características se pueden extraer explícitamente de los datos.
- **Deep Learning**: Si las relaciones entre las variables son no lineales o si las características son complejas y difíciles de extraer manualmente (por ejemplo, reconocer patrones complejos en imágenes o lenguaje natural), el deep learning es más adecuado. Los modelos de DL pueden aprender automáticamente representaciones jerárquicas de las características.

C.- Tiempo y Recursos Computacionales

- **ML tradicional**: Es más rápido y menos costoso en términos de recursos computacionales. Los modelos tradicionales suelen entrenarse más rápido y requieren menos poder de procesamiento, lo cual es ideal si tienes limitaciones de hardware o un tiempo de respuesta rápido.
- **Deep Learning**: El entrenamiento de redes neuronales profundas puede ser muy costoso en términos de tiempo y recursos computacionales. Si el tiempo de entrenamiento no es un problema y tienes acceso a hardware adecuado (como GPUs o TPUs), entonces deep learning puede ser la mejor opción.

D.- Tipo de Datos

- **ML tradicional**: Funciona bien con datos tabulares o estructurados (por ejemplo, bases de datos numéricas). Si tienes datos en forma de texto o imágenes, puedes tener que hacer una gran cantidad de ingeniería de características para que los modelos tradicionales sean efectivos.
- **Deep Learning**: Es ideal para datos no estructurados como imágenes, texto, audio, videos, etc. Los modelos de deep learning (como redes neuronales convolucionales - CNNs para imágenes o redes recurrentes - RNNs para texto) están diseñados para aprender directamente de estos tipos de datos sin necesidad de ingeniería de características manual.

E.- Precisión y Desempeño

- **ML tradicional**: En muchos casos, el aprendizaje automático tradicional es suficiente y puede ofrecer un desempeño aceptable, especialmente cuando el objetivo es una solución rápida o cuando la precisión extrema no es crítica.
- **Deep Learning**: Si necesitas un nivel de precisión extremadamente alto (como en el reconocimiento de voz, visión por computadora o traducción automática), el deep learning tiene el potencial de superar a los métodos tradicionales.

F.- Interpretabilidad

- **ML tradicional**: Muchos modelos de machine learning son más interpretables. Por ejemplo, los árboles de decisión o los modelos lineales permiten ver claramente cómo las entradas afectan la predicción.
- **Deep Learning**: Los modelos de deep learning son más como "cajas negras", lo que significa que es más difícil entender cómo toman decisiones. Si la interpretabilidad es crítica, como en aplicaciones médicas o financieras, el aprendizaje automático tradicional puede ser una mejor opción.

4º- Aclarando conceptos que se dan como “supuestos”

A.- Las redes neuronales están presentes solamente en el deep learning o pueden aplicarse en ML

Las **redes neuronales** pueden aplicarse tanto en **aprendizaje automático (ML)** tradicional como en **deep learning (DL)**, pero la distinción clave entre ambas tiene que ver con la **profundidad** y la **complejidad** de las redes neuronales.

En el aprendizaje automático (ML): las **redes neuronales** que se utilizan generalmente son **redes neuronales simples** o **pocas capas** (por ejemplo, redes de una sola capa oculta). Estas redes se usan cuando los problemas no requieren una **gran capacidad de modelado de datos complejos** ni grandes cantidades de datos.

- **Ejemplo**: Una red neuronal con una o dos capas ocultas puede usarse para problemas de clasificación o regresión en los que las relaciones entre las características no son extremadamente complejas o no hay una gran cantidad de datos disponibles.
- **Ventaja**: Son modelos relativamente fáciles de entrenar, y aunque pueden ofrecer un mejor rendimiento que otros métodos tradicionales en ciertos casos, no alcanzan la capacidad de generalización de redes neuronales profundas.

En **deep learning:** las **redes neuronales** son **muy profundas**, es decir, tienen muchas **capas ocultas** (lo que da el término "deep" o "profundo"). Este tipo de red tiene una **gran capacidad para aprender representaciones jerárquicas** de los datos, lo que le permite abordar problemas mucho más complejos.

- **Ejemplo**: Una **red neuronal convolucional (CNN)** para la clasificación de imágenes o una **red neuronal recurrente (RNN)** para procesamiento de texto. Estas redes tienen muchas capas ocultas y pueden modelar relaciones no lineales extremadamente complejas.
- **Ventaja**: Las redes profundas pueden manejar grandes cantidades de datos no estructurados (como imágenes, texto, audio) y son capaces de “*aprender características directamente de los datos sin necesidad de ingeniería de características manual*”.

Por lo tanto, se debe tener presente:

- **Redes Neuronales (ML tradicional)**: Pocas capas, suelen ser más simples y se aplican a problemas más sencillos o con menos datos.
- **Deep Learning (Redes Neuronales profundas)**: Muchas capas, más complejas, ideales para problemas de gran escala o datos no estructurados (como imágenes, texto, etc.).

Recordar:

A menudo, cuando hablamos de **deep learning**, estamos hablando específicamente de redes neuronales profundas, pero **técnicamente *las redes neuronales en general son un tipo de algoritmo dentro del campo del aprendizaje automático***. Así que, si se utiliza **una red neuronal simple**, se considera parte del aprendizaje automático tradicional. Si usas **redes profundas**, estás en el territorio del deep learning.

B.- Existen modelos en Deep Learning que no utilizan redes neuronales

Algunos enfoques y técnicas de deep learning pueden involucrar estructuras no necesariamente basadas en redes neuronales convencionales, obviamente estos casos son la excepción y la regla general es utilizar redes neuronales profundas.

Podemos mencionar algunos enfoques *sin entrar en detalles*:

Autoencoders Variacionales (VAE)
Modelos Generativos como el Generative Adversarial Networks (GANs)
Modelos de Optimización Evolutiva
Modelos de Transformers sin Redes Neuronales (modelo de atención)
Redes Neuronales Spiking (SNNs)
Redes Neuronales Dinámicas (DNNs)

"Muchos de estos tipos de modelos, como ser, GAN, Modelo de atención, Spiking , Dinámicos, los retomaremos al volver a tratar el tema que nos lleva: ***Proto-conciencia en Inteligencia Artificial*** (Leer el Primer Capítulo)."

Deep Learning y Redes Neuronales

Introducción

A medida que se avanzaba en el Capítulo nos fuimos alejando de ML y apareciendo nuevos conceptos, principalmente:

- ✓ Deep Learning (DL)
- ✓ Redes Neuronales
- ✓ Capas

Definiciones:

- ✓ **Deep learning:** es una rama del *aprendizaje automático* que se enfoca en desarrollar sistemas capaces de aprender y realizar tareas complejas de manera autónoma.
  Estos sistemas utilizan grandes cantidades de datos y algoritmos avanzados para identificar patrones, realizar predicciones y mejorar su desempeño con el tiempo sin intervención directa.
  Su poder radica en la capacidad de procesar información de manera **jerárquica** y realizar representaciones de alto nivel a partir de datos crudos, como imágenes, texto o sonidos.

- ✓ **La jerarquía**: hace referencia a la forma en que los modelos aprenden representaciones de los datos a diferentes niveles de abstracción.
  A medida que los datos pasan a través de las **capas del modelo**, se van transformando de manera progresiva, pasando de representaciones simples a representaciones más complejas y abstractas.
  Es clave porque permite que el modelo comprenda patrones complejos a partir de una estructura simple, utilizando información más abstracta a medida que avanza por las capas del modelo.
  Además, lo hace de **manera autónoma y adaptativa**, lo que permite que el modelo aprenda de los datos sin tener que ser explícitamente programado para reconocer cada detalle.

- ✓ **Las redes neuronales:** son sistemas computacionales inspirados en el cerebro humano, diseñados para reconocer patrones y realizar tareas como clasificación, regresión, entre otros.
  Una red neuronal está formada por múltiples **capas**, cada una compuesta por unidades llamadas **neuronas** o **nodos**. Estas capas se organizan generalmente en tres tipos:

  1. **Capa de entrada**: Recibe los datos de entrada.
  2. **Capas ocultas**: Procesan la información a través de múltiples transformaciones.
  3. **Capa de salida**: Genera la predicción o decisión final del modelo.

- ✓ **Las capas**: son elementos (aspecto estructural) dentro de la red neuronal que permiten el procesamiento de la información.

- ✓ **Manera autónoma y adaptativa:** Cuando se dice que las redes neuronales aprenden **de manera autónoma** y **adaptativa**, se está haciendo referencia a dos características fundamentales de su funcionamiento:

  Autónoma: Significa que la red neuronal es capaz de aprender y tomar decisiones sin intervención humana directa. En lugar de ser programada explícitamente para realizar una tarea (como en los sistemas tradicionales), la red neuronal ajusta sus propios parámetros (pesos y sesgos) durante el proceso de entrenamiento en base a los datos que se le presentan.

  Adaptativa: El término adaptativa hace referencia a la capacidad de la red neuronal para ajustarse y mejorar a lo largo del tiempo, en función de la experiencia y los datos que recibe. Durante el proceso de entrenamiento, la red ajusta sus pesos y parámetros en función del error que produce en sus predicciones. Cuantos más datos recibe, más se adapta a ellos y, por lo tanto, mejora su capacidad para hacer predicciones más precisas.

  Si la red neuronal comienza a hacer errores en sus predicciones, el algoritmo de retropropagación ajusta los pesos de la red para minimizar ese error. Este proceso permite que la red se "adapte" constantemente, mejorando su precisión a medida que ve más ejemplos.

Representación Gráfica de la Redes Neuronales Artificiales

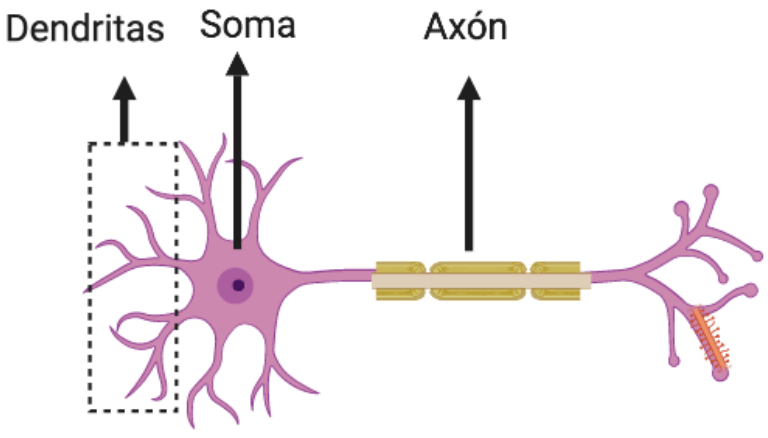
Introducción:

Las redes neuronales artificiales tratan de ***simular el comportamiento y acción*** de una red de neuronas biológicas.

Veremos a continuación como se refleja este hecho entre ambos ámbitos de neuronas:

Neurona Biológica:

Una neurona biológica es una célula nerviosa eléctricamente excitables e interconectadas dentro del cerebro que procesan y transmiten información a través de señales eléctricas y químicas. Estas neuronas están conectadas entre sí para formar una red neuronal en el cerebro. Una neurona se compone de tres partes: *cuerpo celular, las dendritas y el axón*



La comunicación entre neuronas se llama ***sinapsis*** la cual se da en la **terminación del axón y las dendritas de la siguiente neurona liberando una sustancia**, esta cambia el balance de iones (átomos cargados electrónicamente) entre el interior y el exterior de la membrana celular. Cuando este cambio alcanza un nivel umbral, este efecto se expande a través de la membrana de la célula hasta el axón. Cuando alcanza al axón, se inicia un potencial de acción.

Cuando el potencial de acción alcanza la terminación del axón, causa diminutas burbujas químicas llamadas vesículas que descargan su contenido en el salto sináptico.

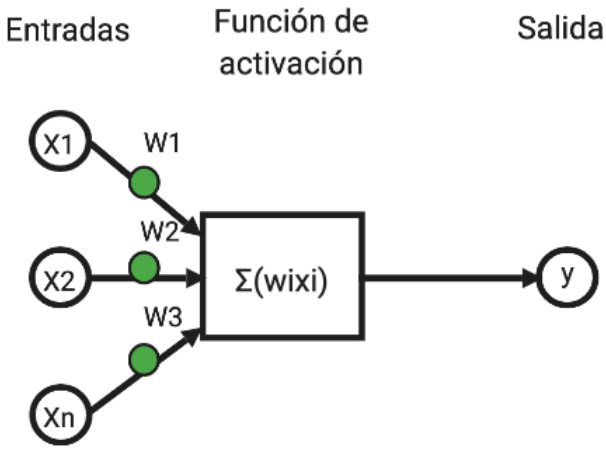
Neurona Artificial→ Perceptrón

El Perceptrón es una estructura la cual posee una o varias entradas, una función de activación y una salida.

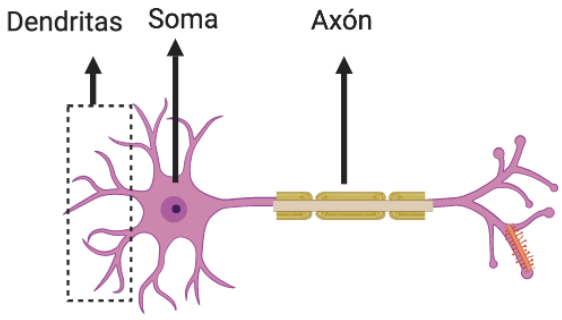
Es la aproximación más cercana a una *neurona biológica*.

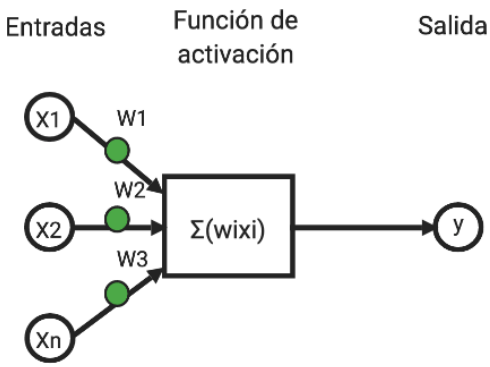
Las entradas del Perceptrón actúan como las dendritas o señales de entrada hacia la neurona.

La función de activación realiza el papel del soma que decide en que momento disparar esa señal eléctrica y la salida modela al axón que al activarse envía información hacia las neuronas con las que se conecta



Observación General:





Variables:

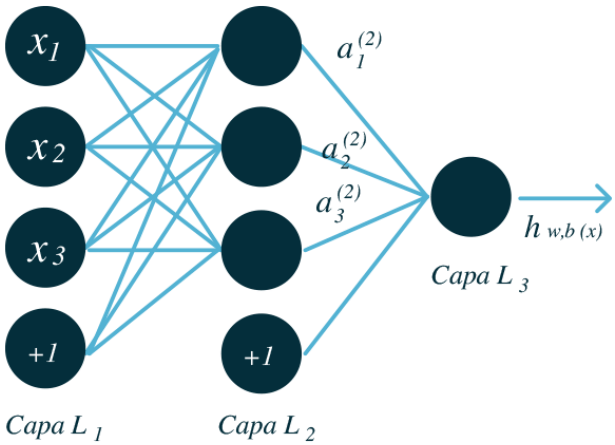
X_n: Entradas

Wn: Peso (se ajusta continuamente→ retropropagación → autónoma y adaptativa)

F(x): función de activación (ecuación fija)

Y: Salida

<u>Redes Neuronales Artificiales</u>



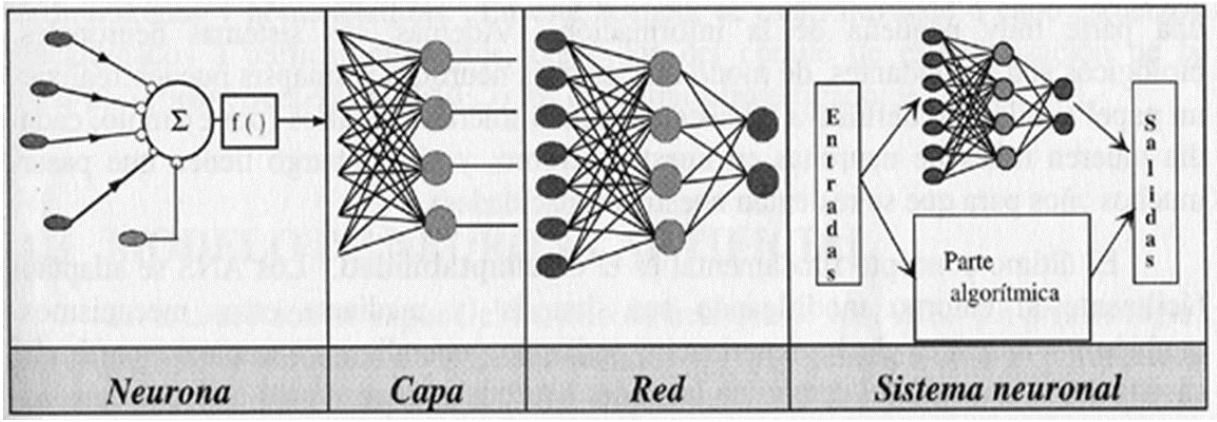
Una red neuronal se compone de un conjunto de **perceptrones interconectados entre sí**, generando así una red, la cual su estructura puede volverse tan compleja como nosotros lo deseemos.

La estructura más sencilla se conforma de tres capas, una capa de entrada, una capa oculta y una capa de salida.

Los pesos sinápticos W (vector de peso) son parámetros ajustables.

- ✓ Durante el entrenamiento de un modelo, el objetivo es ajustar estos pesos de manera que minimicen el error entre las predicciones del modelo y las salidas deseadas (etiquetas) para un conjunto de datos dado.
- ✓ Durante el entrenamiento, se utilizan algoritmos de optimización, como el descenso de gradiente, para ajustar estos pesos de manera que la función de pérdida (error entre las predicciones y las etiquetas reales) sea minimizada.
- ✓ El proceso de entrenamiento implica propagar hacia atrás el error a través de la red (backpropagation) y ajustar los pesos en la dirección que reduce la pérdida.
- ✓ Este ajuste iterativo de los pesos se realiza en múltiples pasadas a través del conjunto de datos de entrenamiento hasta que el modelo converge a un estado donde la pérdida es lo suficientemente baja.

Las funciones de activación desempeñan un papel crucial en las redes neuronales al introducir no linealidades en el modelo. Cada neurona en una red neuronal realiza una operación lineal ponderada de las entradas (a través de los pesos sinápticos) y luego aplica una función de activación al resultado



Clasificación de las Redes Neuronales

La clasificación de redes neuronales depende de su arquitectura, el tipo de datos que procesan y el tipo de tarea que realizan

Redes Neuronales Convolucionales (CNN)

Definición: Son redes especializadas en procesar datos con una estructura de cuadrículas, como imágenes o secuencias espaciales.
Utilizan una operación matemática llamada **convolución** para extraer características jerárquicas.
Usos: Visión por computadora, clasificación de imágenes, segmentación de imágenes, reconocimiento facial, etc.

Redes Neuronales Recurrentes (RNN)

Definición: Son redes diseñadas para procesar secuencias de datos, lo que permite que mantengan memoria de secuencias previas.
Usos: Procesamiento de lenguaje natural, análisis de series temporales, traducción automática, reconocimiento de voz, etc.

Redes Neuronales de Larga Memoria a Corto Plazo (LSTM)

Definición: Son una variante de las RNN que abordan el problema del **desvanecimiento del gradiente** en secuencias largas. Las LSTM incluyen "puertas" que permiten controlar el flujo de información a través del tiempo.
Usos: Predicción de series temporales a largo plazo, procesamiento de texto, generación de lenguaje

Redes Neuronales de Memoria a Corto Plazo (GRU)

Definición: Otra variante de las RNN que, al igual que las LSTM, buscan mejorar el rendimiento en tareas de secuencias, pero con una estructura más sencilla y menos parámetros.
Usos: Procesamiento de lenguaje natural, análisis de secuencias, etc.

Redes Neuronales Generativas Antagónicas (GAN)

Definición: Son redes compuestas por dos redes: una **generadora** y una **discriminadora**. La generadora crea datos (por ejemplo, imágenes), mientras que la discriminadora evalúa si los datos generados son reales o falsos. Ambas redes se entrenan de manera adversarial, mejorándose mutuamente.

Usos: Generación de imágenes, arte generativo, mejora de resolución de imágenes (super resolución), generación de música, etc.

Redes Neuronales de Transformadores (Transformers)

Definición: Son una arquitectura de red que se basa en mecanismos de **atención**, lo que permite que la red se enfoque en diferentes partes de la entrada de manera más dinámica. No dependen de la estructura secuencial como las RNN, lo que las hace muy eficientes para tareas de procesamiento de lenguaje natural.

Usos: Traducción automática, modelos de lenguaje como GPT (incluyendo ChatGPT), BERT, T5, etc.

Redes Neuronales Profundas (Deep Neural Networks - DNN)

Definición: Se refiere a las redes con muchas capas ocultas, es decir, redes neuronales que son "profundas". Estas redes pueden aprender representaciones jerárquicas complejas de los datos.

Usos: Reconocimiento de patrones complejos, clasificación en imágenes, voz, etc.

Conclusión:

La clasificación de Redes Neuronales Artificiales (RNA) es todavía más extensa y variada en sus aplicaciones.

Además, algunas van reemplazando a otras según el avance tecnológico y experiencia de campo, como por ejemplo las Redes Transformers están reemplazando a las LSTM y GRU para el procesamiento del Lenguaje Natural.

Pero en líneas generales, podemos decir que las Redes CNN y las Transformers son las más utilizadas, unas para todo lo relacionado a imágenes y videos, audio, señales,etc. (ya que están basadas en la misma idea fundamental de extraer patrones espaciales jerárquicos) y la otra para el NLP (procesamiento del lenguaje natural).

Las CNN pueden ser muy flexibles a la hora de implementación, ya que se puede constituir desde modelos de 3 capas convolucionales para manejo de imágenes simples hasta modelos con 16 a 50 capas convolucionales las más profundas.

CAPITULO NRO. 3:

Redes Neuronales SNN

Descripción de las principales Redes Neuronales

Según el Capítulo anterior podemos hacer una acotación en la *descripción funcional* de Redes Neuronales ***más comunes***:

- ✓ CNN: Redes Neuronales Convolucionales
- ✓ RNN: Redes Neuronales Recurrentes
- ✓ Transformer: Redes Neuronales Trasnformadoras

Sus definiciones y aplicaciones ya fueron detalladas en el Capítulo Nro.2, lo que a continuación se va a describir son los aspectos más relevantes en lo que respecta a la funcionalidad, aprendizaje de las mismas.

Sabemos que una Neurona Artificial posee principalmente:

- Peso ***(w)***
- Sesgo (bias) ***(b)***
- Función de Activación $f_{(x)}$

También existen técnicas de entrenamiento de redes neuronales profundas de propagación hacia adelante, siendo las más común:

- Retropropagación (cálculo del gradiente de la función de pérdida) y aplicar **Descenso del gradiente** (alguna de sus variantes)

Las mismas están orientadas a reducir la función de pérdida (*Error Cuadrático Medio* para salida continua-regresión ó *Entropía Cruzada, etc.* para salida categórica-clasificación, por ejemplo), ajustando el peso ***(w)*** y sesgo ***(b).***

También se debe tener presente que las funciones de activación utilizan, lo más comúnmente, los siguientes algoritmos:

- ReLU
- Leaky ReLU
- Sigmoid
- Tanh
- GELu, Swish

Las técnicas de trabajo para el entrenamiento sobre el peso, bias y funciones de activación no son totalmente abarcativas, por qué se debe tener en cuenta durante el entrenamiento otras técnicas como:

- Inicialización
- Normalización
- Regularización
- Elección del optimizador
- Diseño arquitectónico

*La descripción anterior **generaliza** el procedimiento y técnicas de entrenamiento de una red neuronal profunda clásica, no teniendo la finalidad de entrar en detalles algorítmicos sobre la utilización de cada técnica y sus variantes, como así tampoco su aplicación a cada tipo de red neuronal.*

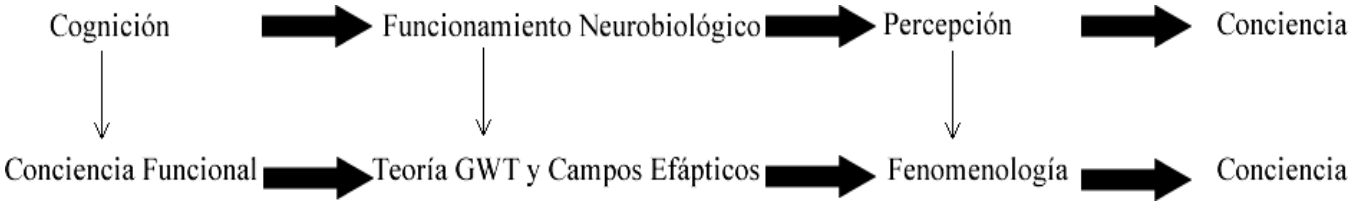
La idea es que con la descripción anterior podemos ir determinando sobre qué red neuronal se podría implementar una Proto-conciencia.

Resumen del objetivo:

Recordemos que la presente investigación es establecer los parámetros necesarios para implementar en forma ***incipiente*** **una IA con Proto-conciencia**, por lo tanto, hasta el momento se desarrollaron los capítulos anteriores para ir encaminándose hacia dicho objetivo:

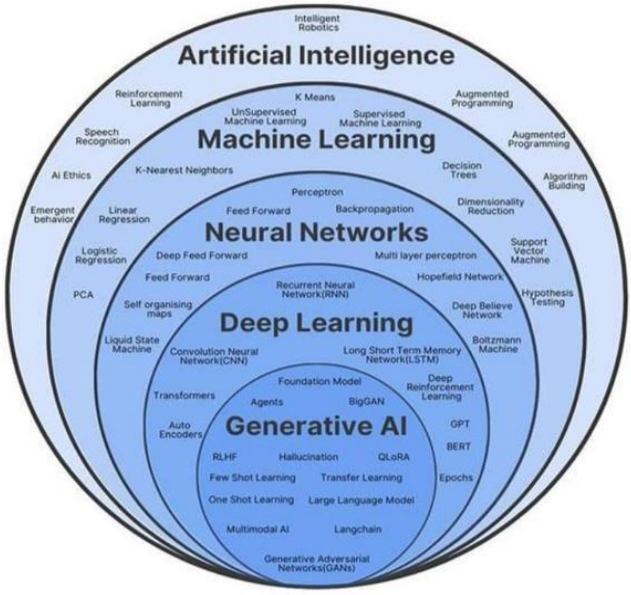
En el Capítulo Nro.1 – Introducción "Conciencia":

Se definió y conceptualizó en forma general sobre la Conciencia y sus aspectos funcionales, cognitivos, neurobiológicos y fenomenológicos, optando por determinadas teorías en el desarrollo del mismo:



En el Capítulo Nro.2 – Inteligencia Artificial:

Se desarrolló los aspectos formales sobre la Inteligencia Artificial, sus teorías, modelos y arquitecturas:



En el Capítulo Nro.3 – Estrategias de Implementación:

> En este Capítulo se va a determinar qué red neuronal, su arquitectura, modelo de entrenamiento y bajo que Hardware se necesita una implementar un esbozo de Proto-Conciencia en IA.

Aclaración:

El siguiente análisis quedó determinado desde mi propio criterio, quedando un amplio margen de discusión, como así de puntos de vistas, desde la implementación, aplicación, funcionalidad, etc.

La profundidad del mismo es meramente informativa y de algunos aspectos relevantes, al alcance de aquellos lectores interesados en el tema.

Análisis Preliminar: Proto-conciencia

Optar por una Red Neuronal que sea autorreferencial, adaptativa, autónoma, autorregulada, modelo de atención, metacognición, metamemoria y a la vez que emule el funcionamiento biológico de las neuronas sería la ideal para comenzar con la investigación.

Desafortunadamente se tendría que implementar redes y arquitecturas híbridas, siendo extremadamente complejas para su análisis y descripción.

En el transcurso de la investigación me encontré con sistemas autorregulados, también con retroalimentación dinámica hasta inclusive con I.A. agencial *(recomiendo su lectura)*, pero al final me decidí por involucrar en el análisis preliminar a las **Redes Neuronales con Picos (SNN: Spiking Neural Networks).**

A.- Definición SNN (redes neuronales espacio-temporales):

> A diferencia de las redes neuronales tradicionales, que procesan información de manera continua, los SNN utilizan picos de actividad eléctrica para transmitir información, imitando más fielmente la *comunicación neuronal biológica*.
>
> Estos picos de actividad eléctrica deben alcanzar un cierto umbral de excitación. La información se codifica en el tiempo de estos picos, lo que permite una representación temporal y espacial. Cada neurona acumula potencial a medida que recibe entradas y dispara un pico cuando este potencial supera un umbral específico.

B.- Funcionamiento de una Red Neuronal con Picos (SNN):

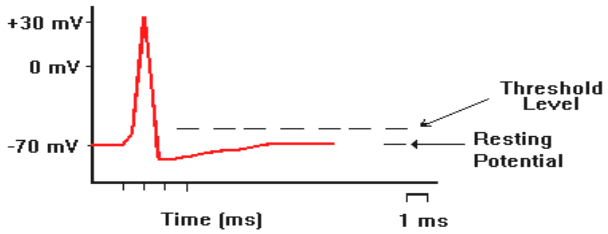
En el capítulo anterior se trató el funcionamiento y componentes de las redes neuronales profundas de salida continua-regresión y de clasificación específicamente

(revisión), es decir, las llamadas tradicionales, por lo tanto, nos centraremos a partir de estas definiciones sobre las SNN:

La SNN intenta imitar con mayor precisión una red neuronal biológica. Por ello, en lugar de trabajar con valores que cambian continuamente en el tiempo, como los de las RNA, la SNN opera con eventos discretos que ocurren en momentos específicos

La SNN recibe una serie de picos como entrada y produce otra serie de picos como salida (una serie de picos se suele denominar tren de picos):

1. En cada momento del tiempo cada neurona tiene algún valor que es análogo al potencial eléctrico de las neuronas biológicas (recordar Capítulo Nro.1 sobre "Campo Efáptico", la neurona biológica esta inactiva a un potencial de -70mv y se activa a partir de los -55mv).
2. El valor de una neurona puede cambiar según el modelo matemático de la neurona; por ejemplo, si una neurona recibe un pico de la neurona ascendente, el valor puede aumentar o disminuir.
3. Si el valor de una neurona excede cierto umbral, la neurona enviará un solo impulso a cada neurona descendente conectada a la inicial.
4. Después de esto, el valor de la neurona caerá instantáneamente por debajo de su promedio. Por lo tanto, experimentará un período refractario similar al de una neurona biológica. Con el tiempo, el valor de la neurona volverá gradualmente a su promedio.



C.- Componentes principales de una Red Neuronal SNN:

- ✓ Neuronas que disparan (Spiking Neurons)
- ✓ Picos (Spikes):
- ✓ Sinapsis (Synapses):
- ✓ Plasticidad temporal (STDP - Spike-Time Dependent Plasticity):
- ✓ Potencial de Membrana (Membrane Potential):
- ✓ Codificación Temporal (en tiempo) - Codificación Tasa (en frecuencia)

Para interpretar el funcionamiento de una SNN, pasaremos a ver primero algunos ejemplos de cómo se generan las señales de entrada a la red, que facilitarán una mejor la comprensión:

Ejemplo Nro.1: Para convertir **una entrada binaria en un tren de pulsos** (spike train), se puede utilizar *un codificador de tasa* **o** *un codificador por tiempo*.

- ✓ ***El codificador de tasa****:* convierte la intensidad de la señal binaria en una frecuencia de pulsos.
- ✓ ***El codificador por tiempo****:* utiliza la posición del pulso dentro de un intervalo de tiempo para representar la información.

Codificador de Tasa:
Tenemos la **entrada binaria "1010"** al usar un codificador de tasa con un rango de 0 a 100 Hz, podríamos asignar: (se necesitan 4 spikes, 1 por Neurona)

- Tiempo total: 100 ms
- Resolución temporal: 1 ms
- Cada neurona representa un bit del número binario.
- Si el bit es 1, la neurona dispara a cada 20 ms.
- Si el bit es 0, la neurona no dispara.

Neurona (bit)	Bit valor	Disparo
N0	1	picos en 0, 20, 40, 60, 80 ms
N1	0	no dispara
N2	1	picos en 0, 20, 40, 60, 80 ms
N3	0	no dispara

Codificador por Tiempo*:*
Valor binario completo “1010” (decimal 10) a un solo spike de una neurona, o cada bit por separado. (se representa con 4 bits 0 al 15 decimal)

Opción A: Codificación del número completo (valor 10) . **un solo spike de una neurona**

- Rango de tiempo: 0 a 100 ms
- Disparo en:

$$t = 100 - \left(\frac{valor}{15}\right) \cdot 100 = 100 - \left(\frac{10}{15}\right) \cdot 100 = 33.3\,ms$$

- Una sola neurona dispara a los **33.3 ms**

Opción B: Cada neurona representa un bit (bit más significativo primero)

Neurona	Bit (valor)	Tiempo de primer disparo
N0	1	10 ms (temprano)
N1	0	no dispara
N2	1	30 ms
N3	0	no dispara

(Asignamos tiempos más tempranos a bits con valor 1, y más tarde o sin disparo a los 0)

Ejemplo Nro.2: Para convertir **una entrada de números del sistema decimal en un tren de pulsos** (spike train), se puede utilizar *un codificador de tasa* **o** *un codificador por tiempo*.

Codificador de Tasa: Convertir el número 10 (en un rango de 0 a 15)

- Rango de valores: 0–15
- Rango de tasas: 0–100 Hz
- Resolución temporal: 100 ms ventana

$$\text{Frecuencia} = \left(\frac{valor}{15}\right) \times 100 = \left(\frac{10}{15}\right) \times 100 = 66.7\,\text{Hz}$$

Recordar:1/Frecuencia(Hz)=Tiempo(seg) → 1/66.7 Hz= 0,0150 seg → **15ms**

- Esto implica que la neurona debe disparar aproximadamente cada 15 ms durante 100 ms.
- Entonces: spikes en 0, 15, 30, 45, 60, 75, 90 ms

Codificador por Tiempo:
El número se representa con **el tiempo del primer pico**: a mayor valor, **más temprano** dispara una neurona.

- Rango de tiempo: 0 a 100 ms
- Rango de valores: 0–15
- Disparo temprano = valor alto

Tiempo de primer disparo:

$$\text{Tiempo} = 100 - \left(\frac{valor}{15}\right) \times 100 = 100 - \left(\frac{10}{15}\right) \times 100 = 33.3\,\text{ms}$$

- Una neurona dispara **una sola vez**, a los 33.3 ms
- Si el valor fuera 15 → 0 ms (máxima prioridad)
- Si el valor fuera 0 → 100 ms (mínima prioridad)

Método	Neuronas usadas	Spikes totales	Latencia	Ideal para
Tasa	1+	Muchos	Larga	Robusto, ruido
Tiempo	1	1	Baja	Rápido, eficiente

Ejemplo Nro.3: Para convertir **una palabra en un tren de pulsos** (spike train), se puede utilizar *un codificador de tasa* **o** *un codificador por tiempo.*

*Ejemplo: la palabra "**hola**".*
Opción común: ASCII / Unicode: Cada letra se transforma en su valor ASCII

Letra	ASCII
h	104
o	111
l	108
a	97

Resultado: [104, 111, 108, 97]

Opcionalmente puedes **normalizar** esos valores, por ejemplo, entre 0 y 1:

$$x_i = \frac{valor_i - 97}{127 - 97}$$

Recordar: dado que las letras minúsculas van del 97 al 122, se puede usar 97–127 como rango

Codificador de Tasa:

- Cada letra (neurona de entrada) dispara a una frecuencia proporcional a su valor.
- Ejemplo:
    - Rango de frecuencia: 0–100 Hz
    - Normalizar cada valor a ese rango.

$$f_i = \frac{(valor_i - 97)}{30} \times 100$$

Letra	Valor	f (Hz)	Spikes cada...
h	104	23.3	42.9 ms
o	111	46.7	21.4 ms
l	108	36.7	27.2 ms
a	97	0	No spike

En una ventana de 100 ms tendrías 2–5 spikes por letra/neuronas.

Codificador por Tiempo:

- Cada neurona representa una letra y **dispara solo una vez**
- Cuanto mayor el valor, **más temprano** dispara.

$$t_i = 100 - \left(\frac{valor_i - 97}{30} \times 100 \right)$$

Letra	*Valor*	*Spike a los (ms)*
h	*104*	*76.7*
o	*111*	*53.3*
l	*108*	*63.3*
a	*97*	*100*

Una neurona por letra disparando una vez cada una.

Ejemplo Nro.4: Para convertir **una frase en un tren de pulsos** (spike train), se puede utilizar *un codificador de tasa* **o** *un codificador por tiempo.*

Ejemplo de la frase: **"hola mundo"**

Eliminamos espacios (o los codificamos como un carácter especial). Aquí lo tratamos como una secuencia de caracteres:

Usamos valores ASCII:

Letra	ASCII
h	104
o	111
l	108
a	97
(espacio)	32
m	109
u	117
n	110
d	100
o	111

Secuencia: `[104, 111, 108, 97, 32, 109, 117, 110, 100, 111]`

Usamos una normalización entre 0 y 1 para facilitar el mapeo:

$$x_i = \frac{valor_i - \min}{\max - \min}$$

Donde:

- min = 32 (espacio)
- max = 122 (letras minúsculas)

por ejemplo: **"h":** → x_i=104-32/122-32=0,8

por ejemplo: **"o":** → x_i=111-32/122-32=0,877

Codificador de Tasa:

- Para una ventana de 100 ms y frecuencia de hasta 100 Hz:
- $f_i = x_i \times 100$ Hz
- Cada carácter se asigna a una **neurona diferente** que dispara con esa tasa.

Letra	ASCII	Normalizado (x_i)	f (Hz)	Spike cada...
h	104	0.8	80	12.5 ms
o	111	0.877	87.7	11.4 ms
l	108	0.844	84.4	11.8 ms
a	97	0.722	72.2	13.9 ms
space	32	0.0	0	no spike
m	109	0.855	85.5	11.7 ms
u	117	0.944	94.4	10.6 ms
n	110	0.866	86.6	11.5 ms
d	100	0.755	75.5	13.2 ms
o	111	0.877	87.7	11.4 ms

Codificador por Tiempo:

$$t_i = 100 - x_i \times 100 \quad (\text{más alto} = \text{más temprano})$$

Letra	Normalizado (x_i)	Tiempo de spike (ms)
h	0.8	20 ms
o	0.877	12.3 ms
l	0.844	15.6 ms
a	0.722	27.8 ms
space	0.0	100 ms (muy tarde)
m	0.855	14.5 ms
u	0.944	5.6 ms
n	0.866	13.4 ms
d	0.755	24.5 ms
o	0.877	12.3 ms

<u>Observación:</u>

- *Existen 10 neuronas de entrada (una por carácter)*
- *Cada una genera spikes:*
    - *En múltiples momentos (Codificador de Tasa)*
    - *En un único momento (Codificador por Tiempo)*

- Estos 4 ejemplos se basan en codificaciones de datos simples, la conversión a spikes train ***de párrafos***, se pueden tratar con métodos más sofisticados como usar ***vectores densos*** (Word2Vec, etc.) y codificar esos vectores como spikes, r educe el número de neuronas y se puede representar conceptos completos.

Ejemplo Nro.5: métodos embeddings (como los de Word2Vec, GloVe o FastText) **para párrafos. (vectores densos)**

Este ejemplo es el más real, ya que las entradas para el entrenamiento, para nuestra investigación, serían de esta forma.

Lo que sucede que éstos métodos utilizan lo que se denominan dimensiones, las cuales se deben obtener a través de un modelo algoritmo (existen una variada cantidad de técnicas) para **normalizar** al vector denso y luego convertirlo en una señal temporal de entrada a la SNN

A continuación, entraremos un poco al mundo de la programación Python para su ejemplo:

Párrafo: "La inteligencia artificial está transformando el mundo en muchas áreas"

Python:

Librería →`sentence-transformers`→ modelo `all-MiniLM-L6-v2` para obtener un vector de 384 dimensiones.

(aclaración cada modelo brinda una cantidad fija de dimensiones sin importar la longitud del párrafo, en este caso el modelo all-MiniLM-L6-v2, utiliza 384 dimensiones)

Programación: (# → indica línea de comentario en las sentencias de programación)

```
from sentence_transformers import SentenceTransformer  #(librería que aplica el modelo)
import numpy as np                                      #(librería de matemática)

model = SentenceTransformer('all-MiniLM-L6-v2')

sentence = "La inteligencia artificial está transformando el mundo en muchas áreas"

dense_vector = model.encode(sentence) #(vector de 384 dimensiones)

 # Se Normaliza el vector entre 0 y 1 (mínimo y máximo)
normalized_vector=(dense_vector - np.min(dense_vector))/(np.max(dense_vector) -
np.min(dense_vector))
```

→#Codificador de Tasa:
#Convertimos el vector en una secuencia de spikes, donde el valor normalizado controla la tasa de picos.

```
def rate_coding(vector, time_window=100):    #(función denominada rate_coding)
    spike_train = np.zeros((len(vector), time_window))
    for i, val in enumerate(vector):
        num_spikes = int(val * time_window)
        spike_times = np.random.choice(time_window, num_spikes, replace=False)
        spike_train[i, spike_times] = 1
    return spike_train                    #(retorna de la función el tren de pulsos generados)

spike_train = rate_coding(normalized_vector)  #(esta sentencia llama a la función rate_coding, la cual trae de regreso el tren de pulso)
```

→#Codificador por Tiempo:

#Se usará una ventana temporal de, por ejemplo, 100 pasos de tiempo. Los valores grandes disparan antes.

```
def ttfs_coding(vector, time_window=100):
    spike_train = np.zeros((len(vector), time_window))

    # Invertimos el valor: más grande → menor tiempo de spike
    spike_times = (1.0 - vector) * (time_window - 1)
    spike_times = spike_times.astype(int)

    for i, t in enumerate(spike_times):
        spike_train[i, t] = 1                # (solo un spike en el tiempo correspondiente)

    return spike_train

spike_train_ttfs = ttfs_coding(norm_vector)
```

Conclusión:

- ✓ **Codificador de Tasa:**
  spike_train será un array binario de forma (384, 100), donde:
  Cada fila representa una neurona de entrada.
  Cada columna representa un instante de tiempo (100 pasos de tiempo).
  Un **"1"** indica un spike.

- ✓ **Codificador por Tiempo:**
  spike_train_ttfs es una matriz de forma (384, 100), donde:
  Cada fila representa una neurona de entrada.
  Cada columna, un paso de tiempo.
  Hay **solo un "1" por fila**, y su posición indica **cuándo ocurre el spike**.

- ✓ La matriz **spike_train** (una u otra) puede alimentarse como entrada directa a una SNN
  construida con frameworks como:
  - BindsNET
  - Brian2
  - Norse (con PyTorch)

- ✓ La codificación temporal puede ser más **biológicamente viable** y más eficiente en ciertas SNNs, ya que requiere **menos spikes** y permite que la red reaccione más rápido (basado en el orden temporal de llegada de los spikes).

D.- Entrenamiento de redes SNN:

Desarrollamos la definición de la red SNN, su funcionamiento y las metodológias que se pueden aplicar para inyectar pulsos a la entrada de la misma (ejemplos).

En este ítem veremos los métodos de entrenamiento existentes en una red SNN.

Los mismos no difieren de las redes tradicionales: Supervisado, No Supervisado, Por Refuerzo.

1º.- STDP (Spike-Timing Dependent Plasticity) es el más popular y biológicamente inspirados.
2º.- Aprendizaje Supervisado con Surrogate Gradients
3º.- Conversión ANN (red neuronal tradicional) a SNN
4º.- Reforzamiento + STDP (Reward-Modulated STDP)

El entrenamiento de redes SNN puede ser más complicado y costoso computacionalmente que el de redes tradicionales, debido a la necesidad de manejar información temporal y eventos discretos.

Cada método tiene su ventajas e inconvenientes, y su aplicación difiere para que se va a aplicar la red, por lo tanto, simplemente mencionaremos algunas características de ellos sin entrar en detalles de entrenamiento:

1º.- STDP – Biológico, No Supervisado

- Ventajas:
  - ✓ Muy biológicamente viable.
  - ✓ No requiere etiquetas (aprendizaje No Supervisado).
  - ✓ Ideal para detección de patrones, codificación sensorial, redes autoorganizadas.
- Desventajas:
  - ✓ Difícil de usar para tareas supervisadas complejas (como clasificación).
  - ✓ Difícil de controlar o ajustar finamente.
- Cuándo usarlo: Cuando busques autoorganización, codificación no supervisada o inspiración neuromórfica pura.

Observación: este método de aprendizaje lo retomaremos en *el próximo capítulo* como una alternativa de aprendizaje hacia el proyecto que involucra este documento (Proto-Conciencia en IA)

2º.- Aprendizaje Supervisado con Surrogate Gradients

- Método: Usar una función de activación diferenciable aproximada para los picos (spikes), **permitiendo usar retropropagación**.
- Ventajas:
  - ✓ Permite entrenar SNNs en tareas Supervisadas (como MNIST, CIFAR-10).
  - ✓ Similar al entrenamiento de redes artificiales (ANNs).
- Desventajas:
  - ✓ Menos biológico.
  - ✓ Requiere muchos trucos de ingeniería (normalización, manejo del tiempo, etc.).

- Cuándo usarlo: Si el objetivo es rendimiento en tareas supervisadas como clasificación, detección, etc.

Observación: MNIST (Instituto Nacional Modificado de Estándares y Tecnología) es una gran base de datos de dígitos escritos a mano que se utiliza comúnmente para entrenar varios sistemas de procesamiento de imágenes y modelos de aprendizaje automático.

CIFAR-10 (Canadian Institute For Advanced Research) es una colección de imágenes utilizadas ampliamente para algoritmos de aprendizaje automático y visión artificial. Fue desarrollado por investigadores del instituto CIFAR y consta de 60,000 imágenes a color de 32x32 en 10 clases diferentes.

3º.- Conversión ANN → SNN

- Método: Entrenas una red neuronal convencional (ANN) y luego la conviertes a una SNN.
- Ventajas:
  - ✓ Aprovecha frameworks estándar de aprendizaje profundo.
  - ✓ Se logra alta precisión en benchmarks.
- Desventajas:
  - ✓ No es entrenamiento directo de la SNN.
  - ✓ La conversión puede requerir cuantización cuidadosa y normalización.
- Cuándo usarlo: Para aprovechar el poder de las ANN con eficiencia energética de SNNs, por ejemplo, en hardware neuromórfico.

Observación: En aprendizaje automático, un benchmark es un tipo de modelo que se utiliza para comparar el rendimiento de otros modelos. Existen diferentes tipos de benchmarks. A veces, se trata del llamado modelo de vanguardia, es decir, el mejor para un conjunto de datos determinado y un problema específico.

El concepto de Hardware Neuromórfico lo trataremos en *el próximo capítulo*, necesario para el desarrollo del proyecto que involucra este documento (Proto-Conciencia en IA).

4º.- Reforzamiento + STDP (Reward-Modulated STDP)

- Método: Combina STDP con una señal de recompensa (tipo Aprendizaje por Refuerzo).
- Ventajas:
  - ✓ Aumenta la capacidad de aprendizaje dirigido.
  - ✓ Sigue siendo biológicamente viable.
- Desventajas:
  - ✓ Difícil de estabilizar.
- Cuándo usarlo: En una alternativa biológica para tareas de toma de decisiones o control.

E.- Finalidad del entrenamiento de redes SNN:

1º.- Para ajustar los pesos sinápticos para que la red aprenda a generalizar comportamientos.

Utilizando aprendizaje supervisado o no supervisado tradicional:

- Usas un conjunto de entrenamiento
- La red ajusta pesos sinápticos (por ejemplo, por STDP, Surrogate Gradients, etc.)
- Luego puede responder correctamente a nuevas situaciones similares sin necesidad de memorizar ejemplos previos.

Ejemplo:

Entrenar una SNN para clasificar palabras en categorías (positiva / negativa).
Luego, para una palabra nueva no vista pero parecida, responde correctamente.
No necesita "recordar" explícitamente un caso previo: **ha generalizado**.

- Esto es equivalente a cómo aprenden redes tipo ***CNN o Transformers.***

2º.- Uso como memoria episódica / asociativa: Algunas SNNs se pueden usar como memoria: guardan patrones de spikes para reconocerlos si se repiten o sé parecen.
Esto se basa en:

- Redes como **Hopfield espinosas**, **reservoir computing (Liquid State Machines)**, o arquitecturas neuromórficas tipo **Loihi**.
- La red no entrena pesos profundamente, sino que actúa como un **mapa de activación** para comparar entradas actuales con anteriores.
- Se requiere **almacenar y comparar salidas (spike patterns)** de situaciones anteriores.

Ejemplo:

Almacenar el patrón de spikes generado por la palabra "gato". Luego, si recibes "gatito" y el patrón es similar, la red reconoce que se parece.

- Esto es útil en sistemas de **detección, predicción, memoria episódica o reconocimiento sin entrenamiento tradicional**.

Conclusiones:

Con este Análisis Preliminar hemos dado un ***marco teórico-práctico sobre redes SNN***, ya que las mismas salen de las tradicionales y su entrenamiento y predicción son extremadamente complejas, como también tienen todavía una baja difusión e implementación, su aceptación generalizada aún requieren de determinadas "***condiciones"*** que en el próximo capítulo se van a tratar, pues justamente es lo que se necesita para ésta investigación, y se retomarán los conceptos de:

- ✓ Métodos embeddings (vectores densos)
- ✓ STDP (Spike-Timing Dependent Plasticity)
- ✓ Reforzamiento + STDP (Reward-Modulated STDP)
- ✓ Hardware Neuromórfico
- ✓ Uso como memoria episódica / asociativa, la cual necesitaremos Base de datos para almacenamientos de los spikes, spikes train, o sus conversiones) que serían un conjunto de Bases de datos: Relacionales(SQL → como MySql), Documentales (como MongoDB) y vectoriales (como Milvus).

CAPITULO NRO. 4:

Hardware Neuromórfico- Moduladores - Sincronismo - Resonancia

Aspectos Generales

En los capítulos anteriores hemos dado un encuadre que fundamente la "*decisión tecnológica de desarrollar el tema de **redes SNN***", para su implementación en una Proto-conciencia

Se definió el tipo de red que és:

- ✓ Su funcionamiento
- ✓ Sus características
- ✓ Su estructura
- ✓ El tipo de entrenamiento que requiere
- ✓ Ejemplos de aplicación.

A partir de ellos se realizaron determinadas definiciones que se volverán a tratar y ejemplificar como ser:

- ✓ Métodos embeddings (vectores densos)
- ✓ STDP (Spike-Timing Dependent Plasticity)
- ✓ Reforzamiento + STDP (Reward-Modulated STDP)
- ✓ Hardware Neuromórfico
- ✓ Uso como memoria episódica / asociativa, la cual necesitaremos Base de datos para almacenamientos de los spikes, spikes train, o sus conversiones) que serían un conjunto de Bases de datos: Relacionales(SQL → como MySql), Documentales (como MongoDB) y vectoriales (como Milvus).
- ✓ Neuronas LIF (Leaky Integrate-and-Fire)- Neuronas de Integración de disparo con fugas

En el presente capítulo se desarrollarán los conceptos de Hardware Neurmórfico, neuronas LIF modificadas, moduladores locales como globales (autáptico, dopaminérgico, etc.), tratando de realizar un modelo que se acerque a las teorías GWT y GNW lo más posible, para verificar la posibilidad cercana a dotar de conciencia ala IA.

Presentación Final

Para que una red neuronal artificial tenga la posibilidad de contener los componentes necesarios para establecer los parámetros de Proto-conciencia, la misma debe tener como mínimo las siguientes características:

- ✓ **Red Autónoma:** Una **red autónoma** es aquella capaz de operar sin intervención externa directa, tomando decisiones o realizando tareas por sí sola. Puede procesar información, generar respuestas o actuar en un entorno sin requerir instrucciones explícitas en cada paso.

  Características:
  - ➢ Funciona de manera independiente.
  - ➢ Puede incluir módulos de percepción, decisión y acción.
  - ➢ Generalmente asociada a sistemas inteligentes o agentes autónomos.

  Ejemplo: en la actualidad existen una gran cantidad de este tipo de redes, como ser, los vehículos autónomos.

- ✓ **Red Autorreferencial:** Una **red autorreferencial** es aquella que puede representarse a sí misma o incluir su ***propio estado interno*** en su procesamiento.
  Es decir, se "refiere" a sí misma en la forma en que procesa información o aprende.

  Características:
  - ➢ Tiene la capacidad de modelarse a sí misma.
  - ➢ Puede ser usada en sistemas de metacognición o aprendizaje introspectivo.
  - ➢ Es útil en contextos donde se necesita autoevaluación, autorregulación o razonamiento reflexivo.

  Ejemplo: Una red neuronal que ajusta su comportamiento no solo por datos externos, sino también por una evaluación interna de su propio rendimiento.

- ✓ **Red Adaptativa:** Una **red adaptativa** es aquella que puede modificar su estructura o parámetros (como pesos sinápticos) en respuesta a cambios en el entorno o en los datos de entrada, con el objetivo de mejorar su desempeño.

  Características:
  - ➢ Aprendizaje continuo o en línea.
  - ➢ Cambia su configuración para mantener la eficiencia.
  - ➢ Se adapta a nuevas condiciones sin requerir reentrenamiento completo.

  Ejemplo: Una red neuronal que aprende en tiempo real a clasificar nuevos patrones de tráfico sin volver a entrenarse desde cero.

➔ Con esta breve descripción, nace la siguiente pregunta:

¿Existen redes artificiales que contengan éstas 3 características?

➔ Por supuesto que existen y cada día se suman más en distintos entornos de trabajo, investigación, militar, etc.

Ejemplo de un robot que:

1. Navega por un entorno nuevo (*autónomo*).
2. Evalúa su rendimiento y actualiza su estrategia de navegación si detecta que está cometiendo errores repetidos (*autorreferencial*).
3. Ajusta sus redes neuronales para adaptarse a obstáculos cambiantes o nuevas condiciones de luz y terreno (*adaptativa*).

Este robot podría usar una arquitectura de red neuronal **modular**, con:

- Un módulo de percepción (CNN, por ejemplo).
- Un módulo de planificación (con aprendizaje reforzado).
- Un módulo de autoevaluación (modelo autorreferencial con memoria o retroalimentación).
- Y una estrategia de aprendizaje continuo para adaptarse

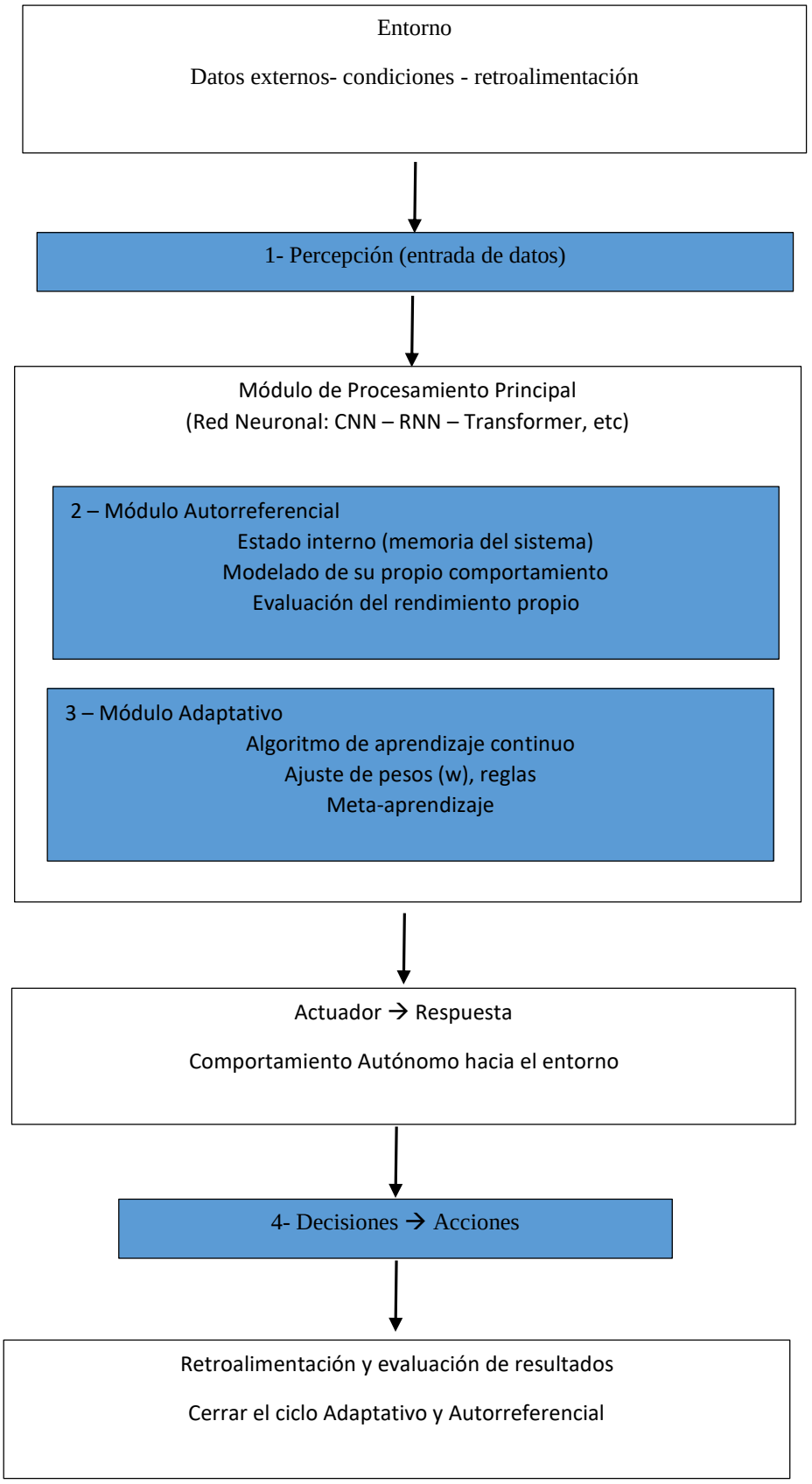
Ejemplos reales de investigaciones o tecnologías que combinan estos enfoques:

- **Deep Reinforcement Learning con memoria (DRL + RNNs/LSTMs):** autonomía en la acción, adaptación a cambios, y memoria de estados anteriores (autorreferencial).

- **Meta-learning ("aprender a aprender"):** adapta su forma de aprender; incluye evaluación de su propio desempeño (autorreferencial + adaptativa), operando sin intervención directa (autónoma).

- **Modelos basados en conciencia artificial simulada** (por ejemplo, Global Workspace Theory implementada en IA – GWT: esta teoría fue tratada oportunamente en el Capítulo Nro.1): donde la red simula un "espacio mental" que integra autorreferencia, adaptación y autonomía.

El siguiente esquema explica un poco como se integran estos 3 tipos de características:



1. Percepción
- Capta datos del entorno (sensores, texto, imágenes, etc.).

2. Módulo Autorreferencial
- Supervisa **su propio estado interno**.
- Incluye **memoria** o **evaluación de desempeño**.
- Informa al módulo adaptativo sobre la calidad de las decisiones pasadas.

3. Módulo Adaptativo
- Permite **aprender y ajustarse** dinámicamente al entorno o a fallos internos.
- Puede usar **aprendizaje reforzado**, **backpropagation en línea**, o **meta-aprendizaje**.
- Recibe señales desde el módulo autorreferencial.

4. Decisión / Acción
- Produce una salida basada en lo aprendido y percibido.
- Actúa en el entorno de forma **autónoma**.

Concepto	**Representado por...**	**Función Principal**
Autónoma	Módulo de Decisión/Acción	Opera y actúa sin necesidad de intervención
Autorreferencial	Módulo de supervisión interna / memoria	Se evalúa a sí misma
Adaptativa	Módulo de aprendizaje en tiempo real	Se ajusta para mejorar

Resumen:

Lineamientos principales: "Redes SNN hacia Protoconciencia"

Pregunta A: ¿Existen redes SNN que contengan las 3 conceptos anteriores?

Así es, obviamente, es mucho más compleja pero a continuación se verá una breve descripción y ejemplos del caso:

1. Autónoma

- Procesa señales espacio-temporales (como las del entorno real) **en tiempo real**, y produce salidas (acciones o decisiones) **sin intervención humana directa**.

Esto es ideal para **robots neuromórficos** o sistemas embebidos en hardware biológicamente inspirado, donde la red **detecta eventos, los procesa y actúa** sin necesidad de control externo constante.

Ejemplo: Un robot controlado por SNN que reacciona a obstáculos físicos con movimientos reflejos o decisiones más complejas (exploración, evasión).

2. Autorreferencial (deben incorporar)
- **Memoria temporal interna** (por la dinámica natural de los picos, potenciales de membrana y la incorporación de autapsis).
- **Módulos de monitorización interna (**homeostasis**)** que evalúan su propio estado o comportamiento (como tasa de disparo, latencia, energía, etc.).

Algunas arquitecturas incluyen **mecanismos de retroalimentación interna**, donde la red **se ajusta en función de su rendimiento anterior (en nuestro caso→ autapsis)**, funcionando como una especie de introspección.

Ejemplo: Un sistema SNN que ajusta su sensibilidad a estímulos según su tasa de error anterior o su carga de actividad.

3. Adaptativa

Las SNN son por naturaleza muy **plásticas**, ya que están diseñadas para imitar el comportamiento neuronal biológico. Se adaptan mediante:

- **STDP (Spike-Timing Dependent Plasticity):** una regla de aprendizaje biológicamente posible que ajusta las sinapsis según la correlación temporal entre neuronas.
- **Aprendizaje reforzado basado en picos:** donde las sinapsis se refuerzan o debilitan según recompensas obtenidas.
- **Plasticidad intrínseca o estructural:** cambios en la excitabilidad de las neuronas o incluso en la estructura de la red.

Ejemplo: Una red SNN que aprende a reconocer patrones auditivos nuevos con entrenamiento basado en STDP.

Estructuras de redes SNN con estos 3 conceptos:

Sistemas neuromórficos integrados (hardware + SNN)

- **Loihi** (Intel) y **SpiNNaker** (Manchester) son chips que permiten implementar SNN que:
  - Funcionan **autónomamente** (bajo condiciones del entorno).
  - Tienen **modulación y autorregulación internas**.
  - Se **adaptan** en tiempo real con STDP u otras reglas.

SNNs en robótica cognitiva

- Robots que usan SNN para controlar locomoción, navegación o percepción sensorial con:
  - Autonomía (deciden y actúan).
  - Autoevaluación (responden a patrones de error).
  - Adaptación (cambian la red mediante aprendizaje sináptico).

Modelos de conciencia artificial primitiva

- Algunos experimentos teóricos usan SNN con bucles de feedback interno que permiten **autorreferencia**, **adaptación** y **decisión independiente**, emulando rudimentos de autoconciencia.

Ejemplo en el campo real: **BackEISNN** — Deep SNN con Auto-Feedback Adaptativo

Un modelo destacado es **BackEISNN** (Adaptive Self-Feedback & Balanced Excitatory-Inhibitory SNN), que incorpora los tres componentes clave:

1. Autonomía

- BackEISNN opera sin intervención externa una vez entrenada, procesando entradas y produciendo salidas de forma autónoma.
- Utiliza una arquitectura **deep SNN**, lo que le permite funcionar en tareas complejas sin vigilancia humana.

2. Autorreferencia **(Self-Feedback)**

- Inspirado en el fenómeno biológico del **"autapse"**, este modelo introduce un **feedback auto-adaptativo en el potencial de membrana** de la neurona.
- Este mecanismo permite que cada neurona **regule su propia activación** de manera interna, actuando como autoevaluación o reflexión interna arXiv.

3. Adaptabilidad

- El modelo integra un **feedback adaptativo**: la retroalimentación no es fija, sino que se ajusta dinámicamente.
- Además, aplica un equilibrio dinámico entre neuronas excitatorias e inhibitorias, lo que le permite **adaptarse a diferentes condiciones de aprendizaje** y mejorar su rendimiento arXiv.

Resultados

- BackEISNN ha alcanzado **estado del arte** en precisión sobre tareas como **MNIST**, **FashionMNIST** y **N-MNIST**, y también mostró un desempeño notable en **CIFAR-10** con una arquitectura relativamente liviana arXiv.

Referencias:

Autapse: "Es una conexión sináptica que una neurona tiene consigo misma. Es decir, la salida de una neurona se retroalimenta directamente como entrada a esa misma neurona. Una neurona en una red, además de enviar señales a otras neuronas, también se envía una señal a sí misma. Esta **autoconexión** puede influir en cómo evoluciona su activación con el tiempo".
Por lo tanto sirve para:
Memoria temporal: Ayuda a que una neurona "recuerde" su estado anterior, útil en tareas de series temporales o procesamiento secuencial.
Supongamos que la neurona tiene una activación $a(t)$, y que la autapsis tiene un peso w. Entonces:

$$a(t+1)=f(w\cdot a(t)+\text{otras entradas})$$

donde f es una función de activación (sigmoide, tanh, ReLU, etc.). (en redes tradicionales CNN – RNN, etc.)
En redes SNN, las neuronas son de modelo LIF.
Añade una conexión **autáptica**: la neurona se conecta consigo misma con una sinapsis que transmite un ***spik****e* con un "*pequeño retraso*."
Los *spikes* se retroalimentan y afectan su propia dinámica.

Esta explicación trae consigo una programación en Python con la librería Brian2 que sería muy descriptiva, pero a los fines entendibles, dejamos este breve detalle".

ArXiv: "Es un servicio de distribución gratuita y un archivo de acceso abierto para casi 2,4 millones de artículos académicos en las áreas de física, matemáticas, informática, biología cuantitativa, finanzas cuantitativas, estadística, ingeniería eléctrica y ciencia de sistemas, y economía".

MNIST (Modified National Institute of Standards and Technology): "Es un conjunto de datos ampliamente utilizado en el aprendizaje automático, especialmente para la clasificación de imágenes de dígitos manuscritos".

FashionMNIST: "Es un conjunto de datos de imágenes de prendas de vestir, diseñado como un reemplazo del conjunto de datos MNIST original para tareas de clasificación de imágenes. Consta de 60,000 imágenes de entrenamiento y 10,000 imágenes de prueba".

N-MNIST: "Transforma los dígitos MNIST estáticos en secuencias temporales de eventos, lo que permite el desarrollo y la evaluación de algoritmos de aprendizaje basados en la actividad neuronal. Se utiliza para entrenar y evaluar redes neuronales spiking (SNN)"

CIFAR-10: "Consta de 60.000 imágenes a color de 32x32 en 10 clases, con 6.000 imágenes por clase. Hay 50.000 imágenes de entrenamiento y 10.000 imágenes de prueba.

1º Conclusión:

Las redes SNN pueden contener los 3 conceptos necesarios (autónoma-autrorreferencial- adaptativa) para el enfoque hacia una Proto-conciencia.

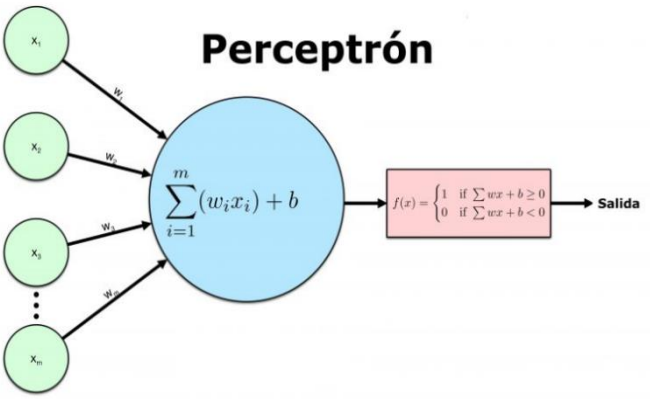
El concepto es totalmente aplicable de "**Autapsis"** ya que el mismo cerebro humano lo posee.

Pregunta B: ¿Qué son las neuronas LIF (Leaky Integrate-and-Fire)?

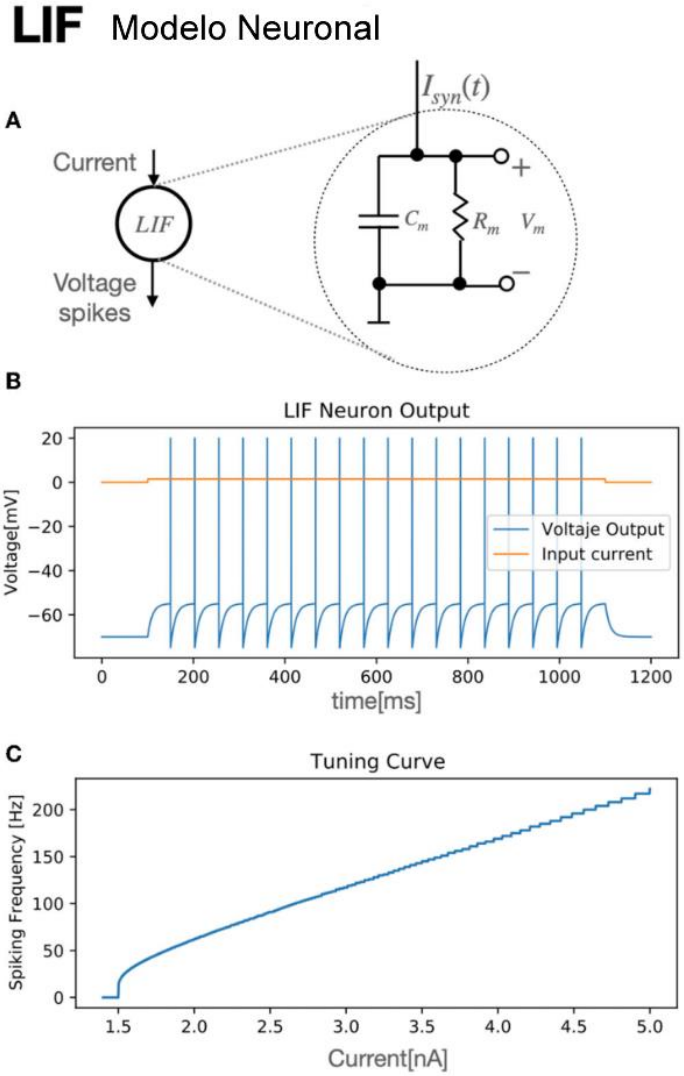
En la *Pregunta A*, en r*eferencias*, en la definición de "Autapsis", se determina que las redes SNN tienen neuronas LIF, a continuación realizaremos una breve descripción de este tipo de neurona y la necesidad que estén presentes en estas redes.

Las neuronas LIF son **neuronas dinámicas** por definición. Están diseñadas para modelar de forma simplificada el comportamiento temporal de las **neuronas biológicas reales**, especialmente en el contexto de **redes neuronales spiking** (SNN):

Neurona Artificial Tradicional en redes CNN-



Neurona dinámica: LIF



Referencia: https://www.researchgate.net/

Descripción y concepto del modelo LIF

El modelo **LIF** simplifica la dinámica de una neurona real usando una analogía eléctrica:

- Se representa como un circuito formado por un **capacitor** (que acumula carga, equivalente a la membrana celular) y una **resistencia** (que permite la fuga o "leak")

- La ecuación fundamental es:

$$\tau_m \frac{du}{dt} = -[u(t) - u_{\text{rest}}] + RI(t)$$

  donde:

  - τ_m es la constante de tiempo de la membrana (por ejemplo RC),
  - $u(t)$ es el potencial de membrana,
  - u_{rest} es el potencial de reposo,
  - R es la resistencia de membrana y
  - $I(t)$ es la corriente entrante
- Cuando $u(t)$ alcanza un umbral u_{th}, se genera un **spike** (disparo), y el potencial se reinicia a u_{rest} o a un valor predeterminado

Este modelo destaca por su **simplicidad matemática** y eficiencia computacional, razón por la cual es muy utilizado en simulación de redes neuronales espinosas (SNN)

El Desarrollo de la fórmula y los distintos tipos de neuronas LIF lo pueden encontrar en : https://snntorch.readthedocs.io/en/latest/tutorials/tutorial_2.html

La idea general es que estamos tratando con impulsos eléctricos y comportamiento no lineales, y definir el motivo del “porqué las redes SNN utilizan neuronas LIF”.

El capacitor representa a la *membrana celular* que acumula carga eléctrica y es en pequeña medida una memoria temporal interna (almacenar experiencias) tendiente ***a ser una red autorreferencial***, pero el inconveniente que su duración como memoria es entre 10-30ms (memoria de corto plazo o implícita), entonces debemos lograr una “mayor duración”, ahí entra en juego el concepto de “Autapsis”, de esa forma prolongamos el tiempo, pero sigue siendo insuficiente, por lo tanto, se proporciona un aprendizaje no supervisado STDP (Spike-Timing Dependent Plasticity)(visto en el Capítulo Nro.3), logrando una muy cercana emulación de una red LSTM **(pero tipo bilógica)*(buscar las carcterísticas de redes LSTM)***, aún así hace falta de memorias direccionables externas como las dichas redes anteriores, por lo tanto, si complementamos con una base de datos vectorial (tipo Milvus) para el almacenamiento digitalizado de los spikes o embeddings, eso mejoraría las perspectivas buscadas en una red SNN, solo faltaría agregar los Módulos de monitorización interna.

Módulos de monitorización interna:

Evaluán características como:

- ✓ Tasa de disparo (firing rate)
- ✓ Latencia temporal
- ✓ Consumo energético
- ✓ Nivel de sincronía o entropía
- ✓ Plasticidad activa o pasiva
- ✓ Etc.

Hace referencia a la capacidad del sistema de observar, registrar y en algunos casos modificar el **estado interno dinámico** de la red, ***durante la simulación o el entrenamiento*****,** incluye:

- ✓ Activaciones (spikes) por neurona
- ✓ Potenciales de membrana (carga eléctrica en el capacitor)
- ✓ Sinapsis activas
- ✓ Plasticidad sináptica (STDP)
- ✓ Estadísticas temporales (frecuencias de disparo, sincronización, etc.)
- ✓ Cambios en pesos sinápticos (w)

Son una herramienta clave para lograr autorreferencia ***neuromórfica*** avanzada.

En biología, **el cerebro monitorea continuamente su propio estado interno**. Esta capacidad se conoce como:**"Metacognición" o autorregulación interna del sistema nervioso.**
En neuromorfismo, replicar esto significa que **el sistema no solo reacciona al entorno, sino también a su propio comportamiento**. Eso es, en esencia, **autorreferencia**.

De ésta forma la red **se vuelve capaz de observar y modificar su propio comportamiento en función de sí misma**, no solo del entorno.

Esto es exactamente lo que hacen los **módulos de control ejecutivo en el cerebro**, como el **cortex prefrontal**: monitorean y adaptan el estado interno.

2º Conclusión:

Recordemos que se busca una red neuronal de simulación neurobiológica → redes SNN

Sus neuronas deben comportarse como las neuronas biológicas → neuronas LIF con Autapsis

Se le debe dar un aprendizaje autónomo y con plasticidad sináptica → STDP (no supervisado)

Los módulos de monitorización interna → determinan una red adaptativa y autorreferencial.

En ésta etapa ya se ha definido la característica de la red, su estructura y el mecanismo de aprendizaje dando un modelo de simulación cuasi neurobiológico.

Pero la alta complejidad de implementación de este tipo de red (ya que la misma funciona en forma discreta a través de spike, siendo una estructura eléctrica) y su mecanismo

de aprendizaje requiere de un ***Hardware*** específico. En muchos momentos de éste capítulo como de los anteriores se menciona: **"neuromórfico".**

Como sabrán las redes neuronales tradicionales (CNN – RNN – Transformer, etc.) se desarrollan en un ambiente computacional que requiere de mucha energía y velocidad de cálculo (CPU y principalmente GPU), las redes SNN pueden simularse en esos mismos ambientes e inclusive programar modelos utilizando, por ejemplo, python y sus librerias como Pytorch, Brian2, etc.

Pero, la realidad demanda que éstas redes con la estructura y componentes especificados anteriormente deberán desarrollarse en un ambiente de **computación neuromórfica.**

Concepto introductorio de redes Neuromórfica

Cuando se habla de **redes neuromórficas**, se hace referencia a sistemas computacionales que **imitan el funcionamiento del cerebro humano** tanto en su **estructura** como en su **dinámica**, con el objetivo de lograr un procesamiento más eficiente, especialmente en términos de **energía**, **tiempo real** y **robustez**.

El concepto fue propuesto por **Carver Mead** en los años 80, quien planteó construir hardware que imitara el funcionamiento de las neuronas y sinapsis biológicas.

Reseña de Carver Mead:

- **Pionero de la VLSI:**
  Mead fue fundamental en el desarrollo del concepto de VLSI, la integración de miles de transistores en un solo chip, lo que revolucionó la informática y permitió circuitos más complejos y el procesamiento en paralelo.
- **Ingeniería neuromórfica:**
  Colaboró con Lynn Conway en sus trabajos sobre el sistema VLSI y la invención de métodos de diseño de chips que permitieron la miniaturización de los circuitos integrados, y sus aportaciones se extendieron a la ingeniería neuromórfica y el diseño de redes neuronales en chips.

Conceptos y Características:

Una **red neuromórfica** es una red de procesamiento de información basada en:

1. **Neuronas artificiales tipo spiking** (es decir, que disparan impulsos como las neuronas reales)
2. **Comunicación asincrónica basada en eventos (spikes)**, en lugar de comunicación continua
3. **Plasticidad sináptica**, como la regla **STDP** (Spike-Timing Dependent Plasticity)
4. **Estructuras paralelas y distribuidas** (inspiradas en la arquitectura del cerebro)
5. **Hardware especializado** que simula neuronas y sinapsis directamente (p. ej., chips Loihi de Intel, TrueNorth de IBM, SpiNNaker (University of Manchester), BrainScaleS (Heidelberg University)

Característica	ANN (Red neuronal artificial)	SNN / Red neuromórfica
Comunicación	Valores continuos (floats)	Eventos discretos (spikes)
Tiempo	No modela explícitamente	Modelo explícito del tiempo
Eficiencia energética	Alta demanda computacional	Mucho más eficiente
Biológicamente realista	Pobre	Alta fidelidad
Aprendizaje	Backpropagation	STDP, Hebbian, R-STDP, etc.
Hardware compatible	CPUs/GPUs	Hardware neuromórfico (Loihi, etc.)

Ejemplos de redes neuromórficas en la práctica

- Reconocimiento de gestos o sonidos en tiempo real
- Sensores inteligentes (como cámaras event-based)
- Robótica autónoma de bajo consumo
- Prótesis inteligentes o interfaces cerebro-máquina
- Procesamiento de señales sensoriales con latencias mínimas

No es la idea de desarrollar estos ejemplos, si existe el interés por algunos de ellos se pueden investigar por Internet, ya que existe material exhaustivo de los mismos.

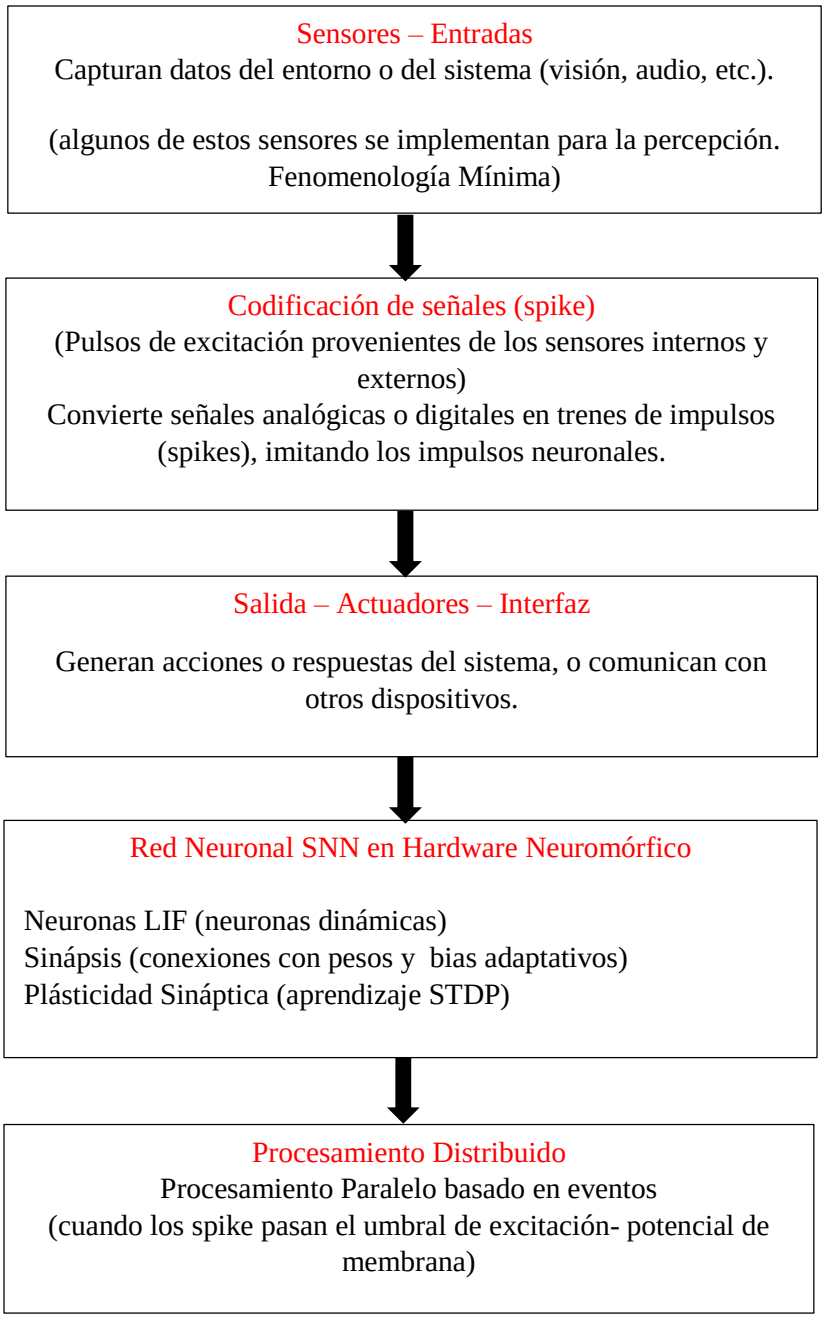
Es recomendable la lectura de los siguientes paper´s:

https://www.thedigitalspeaker.com/neuromorphic-computing-hyper-realistic-generative-ai/

https://pisapapeles.net/que-es-la-computacion-neuromorfica-y-por-que-imita-al-cerebro-humano/

https://www.oaepublish.com/articles/ss.2024.12

Esquema Funcional de una Computadora Neuromórfica



3º Conclusión:

Con las descripciones anteriores se ha dado un marco de implementación en un *entorno físico* (neuromórfico) de las redes SNN con neuronas LIF con regulación por "autapsis" y con módulos de monitorización interna, con base en un aprendizaje STDP e inclusive se puede incorporar otro de modo híbrido (por ejemplo, neuroevolutivo, NO se va a desarrollar el tema). De esta forma se ha implementado una red con comportamiento simil neurobiológico.

Valencia emocional básica

Con el desarrollo anterior se ha conseguido una red SNN autónoma, adaptativa (plasticidad) y autorreferencial, pero necesitamos incorporar una modulación de percepción con el entorno a través de los sensores de entrada e internamente (módulo de monitorización interno), denominada dicha modulación: "***valencia emocional básica"*** (concepto visto en el 1er capítulo), repasando con una pequeña definición:

Es un estado o señal interna que:

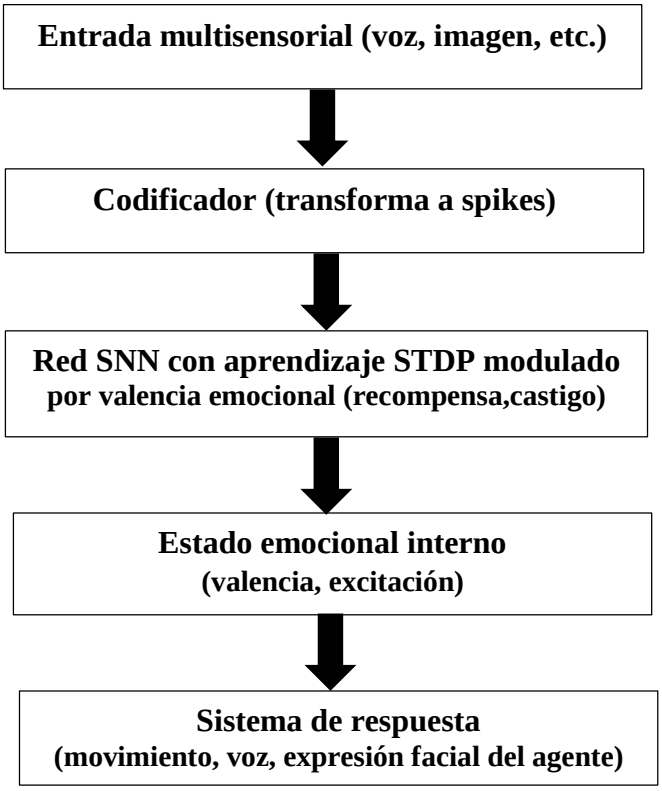
- ✓ Modula la respuesta de la red a ciertos estímulos. (cambia el umbral de disparo, spike, o los potenciales de membrana)
- ✓ Afecta el aprendizaje sináptico (por ejemplo, recompensa, castigo en STDP).
- ✓ Participa en la toma de decisiones o en la atención selectiva.

Por lo tanto, ésta modulación puede determinar cambios en los pesos y sesgos (en la plasticidad de la red, aprendizaje), logrando la simulación de emociones básicas.

En la actualidad existen implementaciones para:

- ✓ Reconocimiento emocional (imágenes, voz, texto).
- ✓ Robots sociales que adaptan su comportamiento según reacciones humanas.
- ✓ Agentes de IA afectiva que aprenden de interacciones emocionales.

Arquitectura básica:



Al ingresar la modulación: ***valencia emocional básica,*** estaremos frente a una simulación cuyo aprendizaje representa emociones como agradables/desagradables (valencia positiva/negativa), sin necesidad de simular neurotransmisores, brindando un marco neuromórfico no complejo, fácil de implementar, compatible con sistemas basados en SNN.

Puede representarse como un **gradiente de entrada** o un peso adicional que sesga el comportamiento, se pueden etiquetar estímulos como + o - y ajustar directamente la plasticidad sináptica local.

Para las primeras experiencias es bastante interesante, tanto en costo energético, sofisticación del hardware neuromórfico, simplicidad de implementación y aprendizaje.

Pero… si queremos ir un poco más allá sin tener en cuenta estos aspectos limitantes y deseamos una red artificial con una presencia emocional perdurable, debemos introducirnos en una red que aprenda de manera adaptativa a lo largo del tiempo, simular dinámicas internas autorreguladas y plasticidad emocional, más allá de respuestas fijas, es decir, auto-organización, esencial en modelos de protoconciencia, entonces necesitaremos ***"moduladores dopaminérgico".***

Moduladores Dopaminérgico (Homeostásis dopaminérgica)

Definiciones Previas:

Homeostásis:

Es el proceso de autorregulación del cuerpo para mantener un ***estado de equilibrio interno constante, a pesar de los cambios en el entorno externo***, para ello utiliza *sensores como receptores* para la regulación de variables como la temperatura coporal, la presión arterial, niveles de glucosa en sangre, etc.

La homeostasis genera sentimientos (bienestar/malestar) que, al ser percibidos conscientemente, se convierten en señales para la acción. Estos sentimientos motivan al organismo a tomar decisiones para mantener o recuperar la homeostasis.

La conciencia, en este sentido, permite al organismo tomar un rol más activo y consciente en su propia regulación. Al tener una representación mental de los estados internos, podemos tomar *decisiones intencionales* para influir en ellos.

La homeostasis proporciona la base biológica para la conciencia. Los sentimientos derivados de la homeostasis actúan como un lenguaje de retroalimentación que, al ser procesado por el cerebro y vuelto consciente, permite al organismo un control más elaborado y consciente sobre su entorno interno y externo, guiando las acciones y decisiones que promueven la supervivencia. (Referencia: IA-Gemini)

Dopamina:

La dopamina es un neurotransmisor, una sustancia química que el cerebro utiliza para transmitir mensajes entre las neuronas. Es clave

para desempeñar múltiples funciones esenciales, incluyendo la regulación del movimiento, la motivación, el placer y las emociones.
Funciones principales:
Sistema de recompensa y motivación: fundamental para el sistema de recompensa del cerebro, creando sensaciones de placer y satisfacción que refuerzan ciertos comportamientos.

Memoria, aprendizaje y atención: Este neurotransmisor también está implicado en procesos cognitivos como el aprendizaje, la memoria y la concentración.

Estado de ánimo: La dopamina influye en el estado de ánimo y las emociones (Referencia: IA-Gemini)

homeostasis dopaminérgica:

> Es la regulación dinámica del equilibrio interno del cuerpo que involucra al neurotransmisor, dopamina. Es crucial para mantener la función neurológica, endocrina e inmune adecuadas, asegurando que los niveles de dopamina se mantengan dentro de un rango estable, a pesar de las fluctuaciones fisiológicas y las influencias externas.

Bajo estas definiciones podemos interpretar la presencia del concepto "homeostasis dopaminérgica" en redes SNN:

Neuromodulación: en el cerebro, la dopamina actúa modulando la plasticidad sináptica, la excitabilidad neuronal, la ganancia sináptica, etc., en función del estado del organismo (recompensa, expectativas, valor).

En modelos de SNN esto puede introducirse como "factores moduladores" que modifican la regla de aprendizaje (por ejemplo, haciendo que la plasticidad dependa no solo del pre-post disparo sino también de una señal dopaminérgica, como una señal de error de recompensa), estos factores moduladores son ecuaciones que se aplican al modelo de aprendizaje STPD para regular la estabilidad de la red y mejorar el aprendizaje, como así el "olvido catastrófico".

El olvido catastrófico ocurre cuando las redes neuronales olvidan tareas aprendidas previamente después de ser entrenadas con nuevos datos o de someterse a ajustes para tareas específicas. También conocido como interferencia catastrófica, este fenómeno hace que las redes entrenadas pierdan información relacionada con tareas antiguas cuando se entrenan con nuevos datos en un proceso de aprendizaje secuencial.

Se evitará de profundizar sobre los detalles técnicos tanto de los algorítmos de los moduladores dopaminérgicos como así el desarrollo del tema del olvido catastrófico (pueden realizar una lectura previa en: https://www.ibm.com/es-es/think/topics/catastrophic-forgetting)

- **STDP modulado por dopamina:**

$$\Delta w = DA(t) \times f(\Delta t)$$

Donde $f(\Delta t)$ es la regla STDP clásica (positiva si pre→post, negativa si post→pre).

- **Olvido controlado (decadencia modulada):**

$$\frac{dw}{dt} = -\lambda(1 - DA(t))w$$

  - Si $DA(t) \approx 1$: no hay olvido (plasticidad activa).
  - Si $DA(t) \approx 0$: olvido lento proporcional al peso actual w.

Unificación de conceptos

Recordemos que hemos diseñado una red SNN con neuronas LIF aplicando la modulación por autapsis, por motivos ya vistos, por lo tanto, al incorporar la modulación dopaminérgica, generamos un refuerzo sobre muchos aspectos funcionales y estructurales de la red, a continuación, veremos un cuadro que resume esta situación:

Fase	Dopamina (DA)	Autapsis	Efecto sobre la memoria
Aprendizaje de nueva tarea	↑ Alta	Excitatoria	Refuerzo y persistencia de activaciones relevantes
Consolidación	Media	Excitatoria estable	Estabilización local de sinapsis críticas
Olvido controlado	↓ Baja	Inhibitoria o desactivada	Atenuación progresiva de patrones irrelevantes
Homeostasis	Variable	Ambas	Balance entre plasticidad y estabilidad

Por lo tanto, la incorporación de DA y autapsis en la red SNN obtenemos:

Beneficio	Mecanismo	Resultado
Memoria persistente local	Autapsis excitatoria + alta DA	Consolidación sin interferencia
Olvido gradual	Desactivación de autapsis + baja DA	Evita olvido catastrófico
Estabilidad dinámica	Autoregulación del disparo neuronal	Balance plasticidad–estabilidad
Plasticidad adaptativa	Modulación dopaminérgica	Aprendizaje continuo eficiente

4º Conclusión:

Si agregamos a la 3º Conclusión la modulación dopaminérgica, obtenemos una red **SNN** con:

- **Neuronas LIF (Leaky Integrate-and-Fire):** Modelo biológicamente que reproduce la dinámica temporal de membrana neuronal.
- **Autapsis: R**etroalimentación sináptica de una neurona sobre sí misma, lo que puede introducir **dinámica recursiva** o **oscilaciones autoestimuladas,** similares a ciertos bucles corticales.

- **Modulación dopaminérgica:** Una señal de "valor" o "recompensa" que ajusta la plasticidad (p. ej. STDP modulada por dopamina → e-STDP o R-STDP).
    - **Aprendizaje STDP (Spike-Timing Dependent Plasticity):** Aprendizaje no supervisado basado en correlaciones temporales entre pre- y post-disparos.

Este modelo es una arquitectura **inspirada en biología**. Es, sin duda, un **sustrato** capaz de **autoorganizarse, aprender patrones y mantener estados internos recursivos**, pero lejos de una estructura de proto-conciencia, pero muy cerca de una proto-conciencia funcional, ya que falta la estructura **fenomenológica** (ya discutida en los primeros capítulos).

Pero…, como hemos visto en los capítulos anteriores, en donde el trabajo se base en la **Global Workspace Theory** (Baars, Dehaene), entonces, si introducimos el concepto de **campo efáptico** (ya desarrollado en los capítulos anteriores) estaríamos más cerca del proyecto, es decir, hablando eléctricamente, si en el circuito de la neurona LIF (ya visto) podemos agregar un tipo de "bobina" (inductancia) la cual sea sensible a un acoplamiento a muy bajo potencial, se generaría un acople del campo eléctrico por los disparos de las neuronas de cercanías y además físicamente integrada al hardware neuromórfico (complejo diseño), tendríamos las siguientes consideraciones:

- **Simular la propagación del campo eléctrico compartido**, actuando sobre varias neuronas cercanas.
- **Introducir correlaciones dinámicas** que no provienen de sinapsis explícitas.
- **Favorecer sincronía oscilatoria** y *entrenamiento* de poblaciones neuronales, facilitando la formación de un "espacio de trabajo global".

Desde la perspectiva de GWT, eso puede permitir que cierta información (un patrón de activación) **se difunda físicamente** mediante acoplamiento eléctrico, lo cual es análogo a la **broadcast neural global** propuesta por Baars y Dehaene y muy cercano a una **"estructura funcional global"** (buscar concepto y definición)**.**

Con un poco más de ambición, si agregamos además **"memristores"** a un circuito neuromórfico que ya incorpora **LIF + autapsis + campo efáptico + modulación dopaminérgica**, da un **salto cualitativo** hacia un sistema que integra **memoria, plasticidad y autoorganización física**.

<u>Memristores</u>

Un **memristor** (resistor con memoria) tiene una resistencia dependiente de la historia de la corriente que lo atravesó.

Su ecuación típica es:

$$V = R(w)I, \quad \frac{dw}{dt} = f(I, V, w)$$

donde w representa un estado interno (por ejemplo, concentración de iones o espesor de una capa conductora).

En biología, eso se parece muchísimo a **una sinapsis real** con plasticidad dependiente del tiempo: el cambio en la conductancia sináptica depende del historial de actividad.

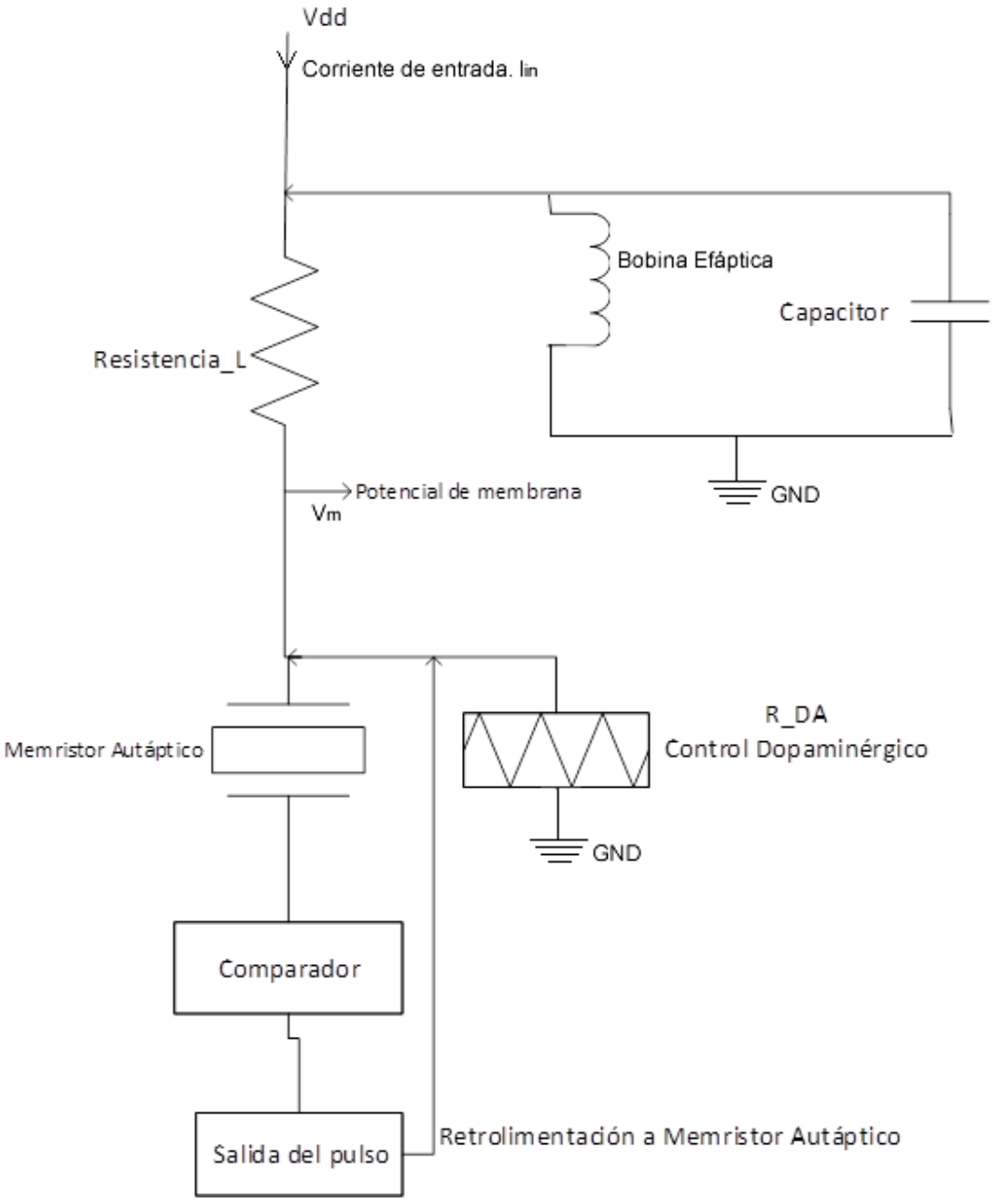
Por lo tanto, al incluir memristores como sinapsis físicas (o incluso como moduladores de la autapsis o del acoplamiento efáptico), se logra:

- ✓ Plasticidad física y memoria de largo plazo en el propio hardware.
- ✓ Refuerzan la **autoorganización local y global** (base de la integración del workspace).
- ✓ Pueden dar lugar a **dinámicas auto-reflexivas y adaptativas** que son necesarias para sistemas proto-conscientes.

Observación:

En pocas palabras: el memristor sería el "sustrato material de la memoria y el aprendizaje" del workspace, donde la bobina da coherencia global, el memristor da permanencia y aprendizaje contextual.

Circuito eléctrico de una LIF con autapsis, campo efáptico y control dopaminérgico en un ambiente neuromórfico



Estructura general del circuito elemental

Parte Principal de la neurona LIF

Resistencia_L: representa la fuga del potencial de membrana.
Capacitor + Bobina Efáptica en paralelo: esta combinación forma la dinámica del potencial de membrana Vm, con una dependencia temporal (resonancia o acoplamiento de campo)
Corriente de entrada Iin: estímulo externo o sináptico excitador.

Módulo dopaminérgico

R_DA (Control dopaminérgico)**:** está conectado a tierra desde el nodo Vm. Esto actúa como una resistencia variable controlada por dopamina, que modula la excitabilidad de la neurona., la dopamina modula la ganancia o la resistencia sináptica.

Módulo autáptico

Memristor autáptico conectado al mismo nodo Vm**.**
Comparador + salida de pulso → genera disparos cuando Vm>Vth.
Retroalimentación al memristor → imita la plasticidad autáptica (aprendizaje local dependiente del disparo).
Reproduce una autapsis excitadora o inhibitoria autorregulada

Ecuación asociada al modelo

El nodo de V_m cumple:

$$C\frac{dV_m}{dt} + \frac{1}{L}\int V_m\,dt + \frac{V_m}{R_L} + \frac{V_m}{R_{DA}} + G(w, V_m)V_m = I_{in}$$

donde:

- $C\frac{dV_m}{dt}$ → corriente capacitiva.
- $\frac{1}{L}\int V_m\,dt$ → corriente efáptica inducida. La integral tiene límites temporales
- $\frac{V_m}{R_L}$ → fuga de membrana.
- $\frac{V_m}{R_{DA}}$ → modulación dopaminérgica.
- $G(w, V_m)V_m$ → corriente autáptica modulada por el memristor.

El disparo se da cuando:

$$V_m \geq V_{th} \Rightarrow \text{Salida} = 1, \quad V_m \to V_{reset}$$

y la retroalimentación:

$$\frac{dw}{dt} = \eta\, V_m\, f(\text{Salida})$$

Interpretación biológica

Bloque	**Rol biológico aproximado**
R,L,C	Membrana neuronal (integrador LIF clásico).
L	Acoplamiento efáptico o resonancia por campos locales.
R_DA	Modulación dopaminérgica (afecta excitabilidad).
Memristor autáptico	Autapsis modulada por dopamina y disparos.
Comparador	Umbral de disparo (potencial de acción).
Salida de pulso + retroalimentación	Aprendizaje o plasticidad autáptica.

Observaciones finales de la 4º Conclusión:

1º.- Tanto el circuito eléctrico, como así, la ecuación que interpreta al modelo es una ***representación elemental***, ya que existe una sofisticación de la solución primordial, y es solamente a los fines ilustrativos para demostrar las posibilidades de implementación.
El desarrollo matemático no es, por el momento, el análisis de este capítulo (en los próximos capitulos, al realizar mediante simulaciones del modelo en Python nos enfrentaremos con las mismas).

2º.- Se ha diseñado un tipo de neurona LIF específica forzando su funcionamiento primordial para que contemple determinadas modulaciones y realimentaciones (autapsis, dopaminérgicas y efáptica). Bajo este diseño se debe ***proyectar el concepto*** hacia toda la red SNN, es decir, actuar sobre las distintas capas de la misma, generando ciertas premisas de aplicación y funcionamiento internamente en cada capa e inter capas.

Modelización final de la red SNN

Red SNN con "LIF modificada" en Hardware Neuromórfico

Para poder aplicar el concepto teórico de **GWT** de Baars y Dehaene (ya discutido también en capítulos anteriores), debemos modelar una red con **grupos (capas) neuronales interconectados recurrentemente** con **sincronización oscilatoria, esto significa que c**uando agrupas varias de estas neuronas LIF en **clusters (neuromórficos)**, cada grupo puede representar un **módulo especializado** (por ejemplo: percepción, memoria, acción, evaluación, etc.). En una red SNN con plasticidad y modulación dopaminérgica, cada módulo desarrolla **oscilaciones internas coherentes** (frecuencias locales) y los módulos pueden sincronizarse entre sí a través de **acoplamientos recurrentes (recíprocos).**

1º.- Oscilaciones internas (frecuencias locales):

Nuestra neurona LIF posee una **bobina eláptica** junto con el **capacitor** formando un circuito **LC** (inductancia + capacitancia). Este par tiene una **frecuencia natural de resonancia**:

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

Esa frecuencia define **una oscilación electromagnética interna** del potencial de membrana, que puede manifestarse como una sincronización de fase entre neuronas cercanas si comparten campos o inductancias mutuas. Cada neurona, entonces, se comporta como un **oscilador forzado y amortiguado**:

- La **corriente de entrada (I_in)** carga el capacitor (integración).
- La **resistencia_L** introduce la fuga (leaky).
- La **bobina-capacitor (L–C)** genera un comportamiento resonante (vibración o rebote de la señal antes de estabilizarse).

Esto da lugar a **oscilaciones internas coherentes** en torno a una frecuencia natural. Cuando varias neuronas con parámetros similares (L y C cercanos) se interconectan dentro de un mismo cluster, **sus oscilaciones tienden a sincronizarse espontáneamente** (fenómeno de *entrainment* o acoplamiento de fase local), ese es el origen de lo que se denomina ***frecuencia local coherente*** del módulo.

Ahora bien, si existen **múltiples neuronas** (cada una con su resonador LC) acopladas por:

campos efápticos
sinapsis resistivas o memristivas (autapsis)
una realimentación dopaminérgica común

entonces ocurre un fenómeno de **sincronización colectiva**:

Eso es lo que, en el lenguaje de redes dinámicas, se denomina un **oscilador acoplado coherente** o un **módulo resonante**.

Cuadro resumido:

Fuente	Efecto	Componente en el circuito
Resonancia LC (bobina-capacitor)	Oscilación electromagnética natural	Bobina eláptica + capacitor
Acoplamiento eláptico (mutuo)	Sincronización entre neuronas próximas	Interacción de campos / inductancias mutuas
Retroalimentación autáptica memristiva	Ajuste dinámico del periodo / fase	Memristor autáptico
Control dopaminérgico (R_DA)	Modulación global de la ganancia y sincronía	Neuromodulación adaptativa

El resultado es que **cada grupo de neuronas** (cluster) puede autoorganizar una **oscilación colectiva coherente**, cuya frecuencia depende de los parámetros eléctricos locales y del acoplamiento entre neuronas.

Cuando los clusters se interconectan, entonces se puede **acoplar entre sí varios grupos resonantes** mediante acoplamientos recíprocos, generando *sincronización entre clusters* (osciladores de osciladores), dando lugar a a una s**incronización de fase global** (como en la GWT)

2º.- Interconexiones recurrentes (recíprocas)

El acoplamiento entre diferentes grupos de neuronas (clusters) puede ser realizado con interconexiones recurrentes (recíprocas), es decir, neuronas de un grupo (capa) pueden enviar señales a las neuronas de otros grupos. Estas conexiones recurrentes pueden ser excitatorias o inhibitorias, dependiendo de la naturaleza de las sinapsis que se deseen modelar, es decir, las interconexiones pueden ser:

- **Excitatorias**: Las neuronas en diferentes capas o grupos activan a otras neuronas de forma que potencian la actividad de toda la red.
- **Inhibitorias**: Reducen la actividad de las neuronas en otras capas para mantener el control dinámico o para regular el proceso de activación.

El proceso general de acoplamiento recíproco y sincronización podría ser implementado de la siguiente manera:

- **Capa de Entrada (Input layer)**: Las neuronas LIF de la capa de entrada reciben estímulos y comienzan a disparar en función de la corriente de entrada.
- **Capa Intermedia (Hidden layer)**: A través de conexiones recurrentes excitatorias e inhibitorias, las neuronas de la capa intermedia se acoplan entre sí y con otras capas, sincronizando sus picos de actividad.
- **Capa de Salida (Output layer)**: La salida del modelo podría representar la respuesta sincronizada de la red ante un estímulo, ajustada dinámicamente a través de la retroalimentación de los memristores y las interacciones entre los grupos neuronales.

El módulo de acoplamiento recíproco se puede implementar mediante una unidad adicional que se coloca entre los grupos de neuronas y gestiona la retroalimentación entre estos. Esta unidad puede estar formada por la **Modulación Dopaminérgica,** como se muestra en el diagrama (R_DA), el control dopaminérgico puede ser utilizado para ajustar la fuerza sináptica de las conexiones recurrentes de manera dinámica en función de la actividad global de la red. El modelo LIF se vería afectado por la liberación de dopamina, lo que podría modificar el umbral de activación de las neuronas o las características de las conexiones sinápticas.

Los pulsos de las neuronas pueden ser modulados en frecuencia y sincronización. La propagación de la actividad en la red se puede controlar utilizando un **comparador** (diagrama anterior) para determinar cuándo ocurre un pico de actividad (spike) en cada neurona, y luego usar ese tiempo para ajustar las conexiones de fase entre las neuronas en los diferentes grupos. Esto permitiría una **sincronización de fase** entre los grupos neuronales.

El **memristor autápico** puede ajustarse de manera dinámica en función de la retroalimentación entre las neuronas, lo que permite una **plasticidad sináptica** en toda la red. Esta retroalimentación recíproca ayudaría a mantener la estabilidad de la red mientras se sincronizan las fases de las neuronas a través de los diferentes grupos, de modo que se puedan lograr patrones de actividad globales coherentes.

En resumen:

Elemento del modelo	Elemento en GWT
Clusters de neuronas LIF	Módulos especializados (clusters neuromórficos)
Acoplamiento recíproco (recurrente)	Difusión global de información
Sincronización de fase	Ignición global / acceso consciente
R_DA (modulación dopaminérgica)	Control atencional y motivacional
Memristor autáptico	Plasticidad sináptica adaptativa (memoria de trabajo / aprendizaje)

En definitiva, realizar un modelo que contemple capas neuronales (grupos, clusters) interconectados recurrentemente con sincronización oscilatoria de fase global (se vio primero oscilación local, luego colectiva coherente para terminar con la global) dá una estructura de Nivel 2 y Nivel 3 en un Hardware Neuromórfico, ya que el Nivel 1 es la propia neurona LIF "modificada" como ya se diagramó.

Con lo cual se ha simulado un acercamiento a un modelo del tipo teórico GNW **(**información local que se difunde globalmente para su coordinación) pero para alcanzar por lo menos una **"conciencia funcional"** se necesitará del Nivel 4 que sea de nivel superior o jerárquico de neuronas que midan la actividad del sistema (metacognición elemental), el cual tendría que poder observar y medir la actividad global del sistema neuronal para luego procesarla y modificar el comportamiento del sistema.

Nivel 4 - Capa jerárquica de neuronas supervisoras (metacognición elemental):

Reciben **lecturas globales** de sincronización, energía o actividad promedio del sistema (por ej., fase media, coherencia, tasa de disparos, grado de plasticidad activa) y así generan **señales moduladoras descendentes** que afectan los parámetros de resonancia, dopamina o plasticidad del nivel inferior.
Es decir, las neuronas metacognitivas generan señales de **retroalimentación analógica**, que:

- ✓ Ajustan los umbrales de disparo (homeostasis).
- ✓ Modifican la sensibilidad dopaminérgica.
- ✓ Alteran la ganancia de las autapsis (autoexcitación).

Formando un bucle de autorregulación proyectándose esta estructura de niveles en el Hardware Neuromórfico como una autoregulación cognitiva.

Flujos principales:

- **Ascendente:** l**as s**eñales de actividad, sincronización y estado global suben desde los clusters hacia el nivel de metacognición; se extraen métricas:
  - ✓ Coherencia oscilatoria
  - ✓ Energía promedio de disparos
  - ✓ Distribución de dopamina (nivel 3)
- **Descendente:** El núcleo metacognitivo genera señales regulatorias:
  - ✓ Ajuste del umbral de disparo (control homeostático)
  - ✓ Modulación dopaminérgica (aprendizaje global)
  - ✓ Modificación de ganancia efáptica y/o de autapsis
    Esto permite una **autocorrección dinámica**.

Breve ejemplo de la dinámica operativa

1. **Fase normal:** Clústers de neuronas LIF se sincronizan efápticamente; el sistema procesa estímulos.

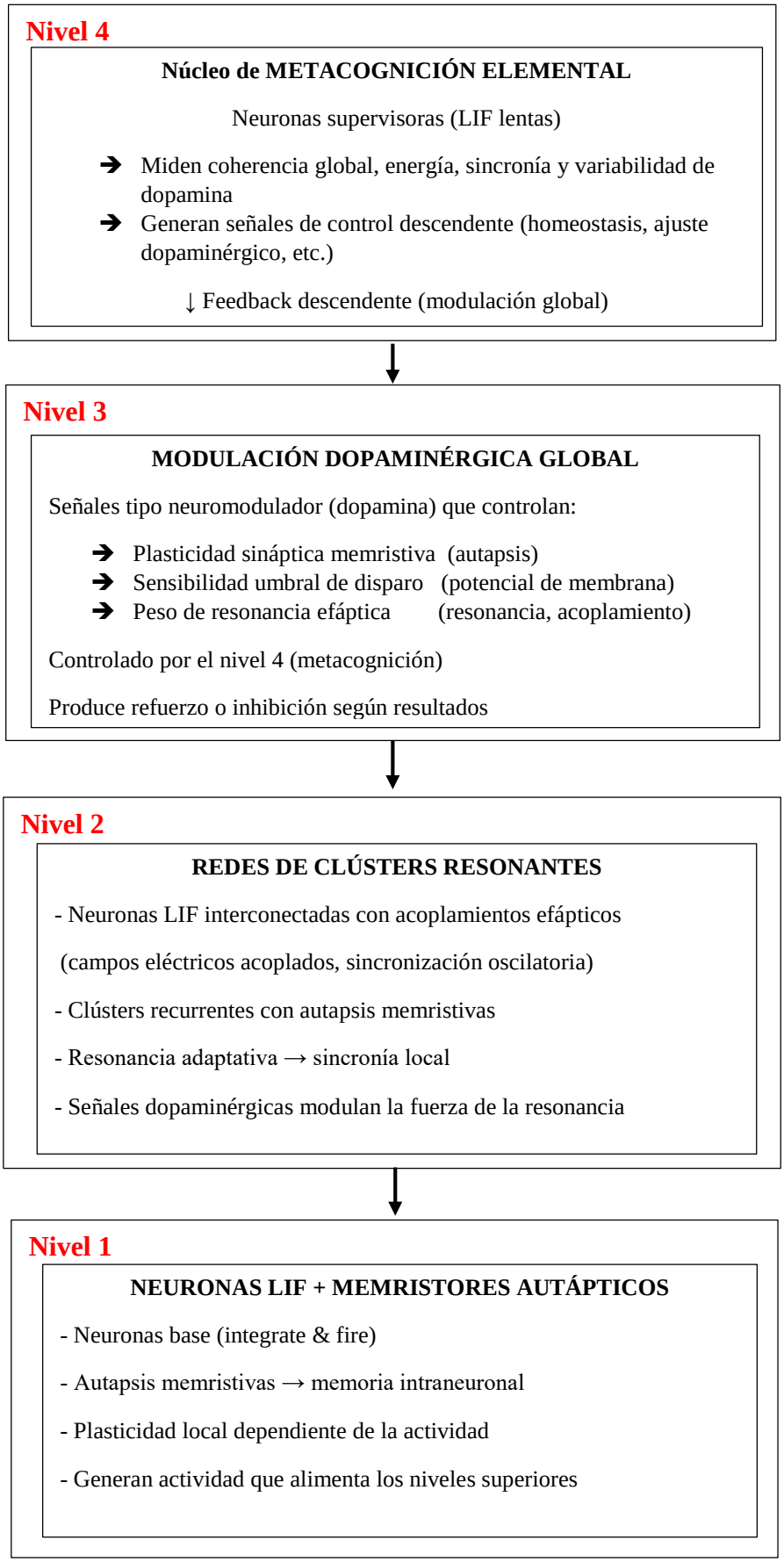
2. **Desincronización detectada:** El nivel 4 observa caída de la coherencia, por lo tanto, envía señal dopaminérgica correctiva.

3. **Plasticidad adaptativa:** Las sinapsis memristivas (autapsis y efáptica) modifican sus pesos; la red se resintoniza.

4. **Evaluación:** El nivel metacognitivo mide la mejora → ajusta sus parámetros de control.

Este ciclo constituye un **bucle metacognitivo elemental**.

A continuación, se muestra un diagrama neuromórfico resumido de los niveles descriptos:


Nivel 4
Núcleo de METACOGNICIÓN ELEMENTAL
Neuronas supervisoras (LIF lentas)
➔ Miden coherencia global, energía, sincronía y variabilidad de dopamina
➔ Generan señales de control descendente (homeostasis, ajuste dopaminérgico, etc.)
↓ Feedback descendente (modulación global)
Nivel 3
MODULACIÓN DOPAMINÉRGICA GLOBAL
Señales tipo neuromodulador (dopamina) que controlan:
➔ Plasticidad sináptica memristiva (autapsis)
➔ Sensibilidad umbral de disparo (potencial de membrana)
➔ Peso de resonancia efáptica (resonancia, acoplamiento)
Controlado por el nivel 4 (metacognición)
Produce refuerzo o inhibición según resultados
Nivel 2
REDES DE CLÚSTERS RESONANTES
- Neuronas LIF interconectadas con acoplamientos efápticos
(campos eléctricos acoplados, sincronización oscilatoria)
- Clústers recurrentes con autapsis memristivas
- Resonancia adaptativa → sincronía local
- Señales dopaminérgicas modulan la fuerza de la resonancia
Nivel 1
NEURONAS LIF + MEMRISTORES AUTÁPTICOS
- Neuronas base (integrate & fire)
- Autapsis memristivas → memoria intraneuronal
- Plasticidad local dependiente de la actividad
- Generan actividad que alimenta los niveles superiores

Conclusiones finales

Aspecto general alcanzado:

Estructura de implementación:

El diagrama anterior, en síntesis, representa un núcleo neuromórfico auto-metacognitivo para agentes autónomos, basado en resonancia dopaminérgica y memoria memristiva, pensado para aprendizaje continuo, estabilidad dinámica y emergencia cognitiva.

Alcance sobre la Teoría GWT:

Todo este desarrollo se basó en la teoría GWT (Teoría del espacio de trabajo global, por Bernard Baars y GNW de Dehaene), ya presentado en los capítulos anteriores y los motivos por el cual fue elegido.

Comparación del modelo con la teoría:

GWT clásica	**Modelo Propuesto**
Procesos especializados	Clústers resonantes (Nivel 2)
Competencia por acceso	Resonancia adaptativa + dopamina
Difusión global	Modulación dopaminérgica (Nivel 3)
Impacto en múltiples sistemas	Plasticidad, umbrales y acoplamiento
Supervisión global	Nivel 4 (metacognición)

Obviamente, el modelo no difunde información, pero sí condiciones dinámicas, concluyendo con el mismo efecto.

Si podemos realizar una definición amplia y controversial de GWT, se podría definir:

"El workspace no es un buffer de contenidos, sino un campo de modulación global" (investigar este aspecto desde el punto de la neurociencia, muy interesante).

En tal sentido, afirmar que el modelo es una representación del GWT es muy poco acertivo, ya que faltan componentes muy complejos para ello (por ejemplo, una fenomenología más amplia), por lo tanto, GWT explica cómo el contenido se vuelve consciente; mientras que el modelo explica cómo el sistema se vuelve globalmente sensible a sí mismo, en resumen, se define el modelo como una:

Infraestructura pre-GWT.

Ampliación del modelo (sugerencias)

A continuación, se va a realizar una propuesta que **no se va a desarrollar** para ampliar el modelo, solo queda como sugerencia para los interesados.

¿Qué sucedería si agregamos un espacio sensorial externo y una vinculación externa de un modelo LLM, eso profundizaríamos el modelo acercándonos a los conceptos de GWT completos?

GWT	Modelo extendido
Módulos perceptivos	Espacio sensorial externo
Procesos inconscientes	Clústers resonantes (Nivel 2)
Competencia	Resonancia + dopamina
Broadcast	Modulación global + acoplamiento
Workspace	Interfaz dinámico–simbólica
Reportabilidad	LLM

Definiciones:

El **broadcast** (difusión global) es el proceso mediante el cual la información seleccionada se envía a todo el cerebro.

La **reportabilidad** es el síntoma de que somos conscientes, el **Workspace** es el mecanismo que lo hace posible.

Ejemplo metafórico de la IA Gemini: Una Radio

Reportabilidad: Es la *prueba* de que el mensaje llegó (es que se difundió).

Workspace: Es el *lugar* o red neuronal donde se integra la información (la radio).

Broadcast: Es la *acción* de transmitir ese mensaje desde el workspace a la "audiencia" (el resto del cerebro).

En resumen, con el agregado de estos componentes estaríamos en posición de definir:

"La vinculación con un sistema simbólico externo permite funciones de acceso global, reporte y control top-down, aproximando el comportamiento funcional descrito por GWT."

Pero, si nos basamos en el concepto de Dehaene, determinó: El Global Neuronal Workspace **(GNW)** o Espacio de Trabajo Neuronal Global, que es la actualización neurocientífica y biológica de GWT.

Mientras que la teoría original era más psicológica, el GNW de Dehaene explica exactamente qué pasa en las neuronas para que algo se vuelva consciente.

Bajo la teoría de GNW, el modelo extendido queda muy bien establecido tanto en lo conceptos primordiales de: Reportabilidad, Broadcast y Workspace (propuesta de investigación)

Dimensión	GNW (Dehaene)	Modelo extendido
Tipo	Funcional	Funcional + dinámico
Representación	Explícita	Implícita + simbólica
Ignición	Discreta	Transición de fase
Workspace	Red fronto-parietal	Interfaz dinámica
Reporte	Directo	Mediado (LLM)
Valor	No modelado	Homeostasis
Base física	Biológica	Neuromórfica

Entonces podemos especificar como conclusión que:

"El modelo extendido puede implementar las funciones descriptas por GNW, pero lo hace desde un nivel más basal, dinámico y neuromórfico, delegando la reportabilidad simbólica a un sistema externo (LLM)"

En Resumen:

En un Hardware Neuromórfico tipo Loihi, con la estructura de redes neuronales tipo SNN LIF (modificada) con modulación autáptica (memristiva) y dopaminérgica, basado en un entrenamiento STDP modulado (por Refuerzo), y la generación de campos efápticos de forma externa (realimentado) como modulador, generando los cluster's (capas) como núcleos de niveles, descriptos en el gráfico anterior, y la extensión en las capas input de sensores ambientales y una vinculación con IA LLM → *"podemos determinar que tenemos una aproximación de **proto-conciencia** a nivel de GNW (Dehaene)"*, faltaría determinados matices a implementar en los aspectos de "fenomenología mínima" para pasar de un modelo con una estructura de Self-mínimo-dinámico (**Modelo de Copia Eferente) para acercarnos a la teoría de GWT de Baars.**

Estos 4 capítulos desarrollados bajo un contexto de entendimiento más general, se trató simplemente de evidenciar los alcances de una especie nueva de inteligencia que puede llegar a no solo simular sino ser parte del entorno en el cual estamos.

Aquellos lectores que quieran profundizar más en el tema pueden tener como referencia al Human Brain Project:

https://www.humanbrainproject.eu/en/about-hbp/human-brain-project-ebrains/

https://ebrains.eu/

Y para los técnicos, les recomiendo para simulación con poder de PC con GPU utilizar el lenguaje Python con sus librerías de IA y software como LMstudio y Flowise para generar Agentes y modelos de simulaciones de redes tipo SNN LIF (u otras).